THIS WEEK

A watchdog with bite

The world must strengthen the ability of the International Atomic Energy Agency to make independent assessments of nuclear safety.

In a recent press conference at the International Atomic Energy Agency (IAEA) in Vienna, a reporter asked a simple question. Chronicling the ongoing nuclear emergency at the Fukushima Daiichi nuclear power plant in Japan, the agency's website consistently referred to "white smoke" rising from the reactors. Why, the journalist asked, did the agency put quotations around the words white smoke?

Denis Flory, the agency's head of nuclear safety and security, said the term arose from lengthy discussions with Japan's nuclear regulator over how to translate the phrase "白い湯気のような煙" (*shiroi yuge noyouna kemuri*), the words used in official Japanese statements. "We got the answer that it meant 'white smoke', so this is why we use 'white smoke,'" he said flatly.

Even by the strict standards of international organizations, the IAEA chooses its words carefully. As the globe's nuclear watchdog, it must simultaneously pronounce on a nation's nuclear programme while being careful not to accuse the country of wanting to develop weapons. Its statements are sometimes cryptic, but they are vital for upholding the delicate Nuclear Non-Proliferation Treaty, which is designed to halt the spread of nuclear weapons.

In the latest nuclear emergency in Japan, however, the IAEA's agonizing over its choice of words has not helped to allay public fears or clarify the situation at the reactors. As illustrated by its derivative use of the term 'white smoke', the agency has been reluctant to deviate even slightly from information delivered by the Japanese government. Its press conferences have been rapid-fire deliveries of temperatures, pressures and radiation readings handed to them by government sources, often with little context.

The agency has good reason to avoid annoying Japan, which is one of 35 members of the board of governors that oversees the IAEA and its budget. Because of the security role played by the organization, these nations have kept the IAEA on a short leash. In the area of nuclear safety, even the rating of a nuclear emergency is out of its hands: individual nations, not the IAEA, judge the severity of an accident.

Nuclear accidents are politically and commercially sensitive events, and it is understandable that countries do not want to cede control of their management to an international body. And nor should they: plant operators are often the best qualified to handle an emergency, and nations must take the responsibility for protecting their citizens.

Yet these nations, and the public at large, would be better served by an IAEA more able to deliver frank and independent assessments of nuclear crises as they unfold. In the aftermath of Fukushima, statements from the Japanese government were often confused. It initially rated the event as an "accident with wider consequences", and then upgraded it to a Chernobyl-scale event a month later, raising anxiety across the country. Moreover, far more severe assessments consistently came from others on the ground, notably the US Nuclear Regulatory Commission. An impartial and authoritative international voice would have been invaluable to avoid at least some of this confusion.

Next month, the IAEA will hold a conference of ministers to discuss lessons to be learned from the Fukushima accident (see page 397). The countries should give the IAEA an explicit mandate, and the necessary resources, to deliver its own safety assessments, both in times of crisis and during the normal operation of nuclear power plants. This more active role would be extremely sensitive, but the IAEA is up to the task. In its job as a nuclear watchdog, the agency already employs highly trained inspectors who regularly visit commercial power plants. The remote systems it uses to monitor nuclear materials could be extended to automatically report conditions at a plant during an emergency. Most important, the agency is politically savvy enough to avoid embarrassing its member states, unless absolutely necessary.

> *"The public would be better served by an IAEA more able to deliver frank and independent assessments of nuclear crises as they unfold."*

In the case of Fukushima Daiichi, an IAEA acting in this way might have strengthened the Japanese position. Japan was criticized in the first days of the crisis for providing too little information on conditions at the plant. An IAEA assessment, based on independent data, could have provided backing for the Japanese decision to rapidly evacuate the surrounding population. It could have provided some reassurance to a panicked population that the government knew what it was doing.

As long as there is nuclear power, there will be the risk of a nuclear emergency. Giving the IAEA the rights and means to pursue a safety agenda cannot prevent such events, but it can reduce their likelihood and strengthen the world's response. ∎

A united front

Pharmaceutical firms should come clean to tackle drug contamination.

When biotechnology company Genzyme announced the presence of a contaminating virus at its drug-manufacturing plant in Allston, Massachusetts, in 2009, patients were told not to worry. Only a small stockpile of uncontaminated drugs existed, but the company said that it would resume production within two months.

Two years and a host of manufacturing problems later, Genzyme still cannot supply enough of its treatment for Fabry's disease, a rare and potentially lethal enzyme deficiency. Genzyme's replacement-enzyme drug, Fabrazyme, which is made at the Allston plant, has been rationed since 2009 so that patients receive smaller doses than initially

recommended. And those diagnosed with the disease after rationing began are barred from receiving Fabrazyme. The restrictions understandably make patients uneasy: many see their symptoms worsening under the new dose regime, and some have started a lawsuit against the firm. In 2010, the European Medicines Agency reported that adverse events in patients with Fabry's disease had risen since the shortage, and advised doctors to prescribe the full dose again. Genzyme's stock price dived amid screams from investors, and the company agreed early this year to be acquired by Paris-based pharmaceutical company Sanofi-aventis.

The fiasco sounded alarm bells across an industry familiar with the difficulty of manufacturing biological molecules such as antibodies and enzymes for use as drugs. These 'biologics' were once the domain of speciality biotechnology firms, but are now being produced in large quantities. The number of clinical trials involving a biologic increased from 1,197 between 2000 and 2005 to almost 6,000 in the following five years. And in 2010, the drugs brought in US$40 billion in sales worldwide.

But success has its price. Unlike the manufacture of small-molecule drugs, which typically relies purely on large-scale chemical synthesis, biomanufacturing usually involves massive cultures of live cells maintained in rich, contamination-prone media. Anyone who has struggled to keep a 1-litre laboratory cell culture sterile will appreciate the challenge of doing the same for a 10,000-litre reactor. Viruses are stealthy intruders and can lie undetected in a culture for weeks, while the infected cells move down the pipeline to spread the scourge through the manufacturing facility — into those 10,000-litre reactors and through million-dollar chromatography columns.

This means that viral contamination can shut down drug production for months and cost a company millions of dollars, interrupting drug supplies and leaving patients vulnerable.

At least 17 incidences of viral contamination in biologics have been reported, but industry insiders say that many more go unreported. Rather than risk negative publicity and lawsuits, companies have largely chosen to keep the details of contamination, and even their occurrence, secret — even, at times, from government regulators. Genzyme's experience, which legally had to be made public because it caused a significant drug shortage, may have only deepened industry's fears of going public.

"Viruses are stealthy intruders and can lie undetected in a culture for weeks."

But although secrecy may make short-term business sense, it hampers industry's collective ability to learn from these catastrophes.

Down the road from Genzyme's troubled plant, researchers at the Massachusetts Institute of Technology in Cambridge are forming a consortium with industry to tackle the problem. The academic organizers hope that it will encourage greater openness and allow industry partners to divulge confidential information under the protection of non-disclosure agreements. The consortium plans to draw lessons from contamination data that could benefit the industry as a whole, and to publish answers to questions such as where viral contaminants originate and what the best way to detect and eradicate them is, or how to prevent them altogether.

These are crucial questions as interest grows in lucrative biopharmaceuticals. The answers should be able to guide research to reduce the incidence and impact of viral contamination on drug manufacturing.

At present, only six companies have signed up to participate in the study's pilot phase, and the organizers say they will eventually need at least twenty more to draw meaningful conclusions. More biotechnology companies should embrace this rare and valuable opportunity to pool resources without compromising their business interests. It could benefit patients and investors alike. ∎

Seeing REDD

Plans to conserve the world's tropical forests must respect the rights of indigenous peoples.

As 'REDD' projects to protect forests in developing countries gain pace, campaigners and other groups representing indigenous peoples have warned that the plans could offer little benefit to local communities that depend on the forests for their livelihoods.

REDD — reducing emissions from deforestation and forest degradation — is touted by proponents as win–win for both conservation and poverty reduction. It is based on taking money from polluters in the developed world and channelling it to tropical nations for use in protection of carbon stocks. The agreement that covers such projects, signed at the United Nations climate meeting in Cancún, Mexico, last year, includes environmental and social safeguards that call for respect for the rights of local and indigenous peoples. But forest-dependent communities and human-rights organizations fear that these provisions offer weak and ineffective protection.

These concerns are starting to play out on the ground. A study by UK-based human-rights group the Forest Peoples Programme (FPP), which looked at nine REDD pilot projects in Cameroon, warns that forest communities there have not been adequately consulted on efforts to move on from the pilot schemes to develop national REDD plans. In addition, the national plans include no measures to protect the rights of these people — such as seeking their free, prior and informed consent to projects that may affect them — nor to ensure that they benefit.

REDD was always going to have teething problems, and there will be opportunities to address these concerns. Eyes are already on an upcoming meeting of the Forest Carbon Partnership Facility (FCPF) — a global fund administered by the World Bank to help developing nations to devise national REDD plans. At the meeting on 20–22 June in Oslo, Cameroon will present its plans, and will ask for up to US$3.6 million to start implementing them.

Will attending conservation organizations such as the WWF, which led the development of Cameroon's REDD plans, have time to note and attempt to rectify the shortcomings identified by the FPP in time for the meeting? Perhaps not, but a subsequent meeting of scientists, international organizations and donors to discuss the social sustainability of REDD will certainly have the opportunity to examine them. The Oslo REDD Exchange will take place on 23–24 June.

An important first step would be for organizations involved in funding and driving REDD projects, such as the World Bank, to take the involvement of local communities more seriously. The FCPF has yet to finalize standards and safeguards for activities it funds, for example, those governing human rights. And it remains unclear what standards REDD projects will be measured against, given that the FCPF is just one of a number of donors. Until these issues are resolved, it will be impossible to tell whether adequate precautions are in place.

Many who follow these issues closely argue that the World Bank must lead by example, and could start by bolstering its own policies on the rights of indigenous peoples. Currently, the bank requires indigenous peoples to be 'consulted' on funded projects that may affect them. Human-rights campaigners would like to see this provision strengthened so that 'consent' is required. They are hoping that this will be a key feature of a review the bank launched last month to examine its operational safeguard policies.

The REDD initiative is too important to be undermined by a reckless disregard for indigenous peoples. It is vital that key players use this year's opportunities to steer it back on course. To ensure that projects on the ground run straight, funders must set a good example. Otherwise, a major opportunity to reduce carbon emissions and improve people's livelihoods will fail before it has a chance to succeed. ∎

↻ NATURE.COM
To comment online, click on Editorials at:
go.nature.com/xhunqv

End the wasteful tyranny of reviewer experiments

Peer review of scientific papers in top journals is bogged down by unnecessary demands for extra lab work, argues **Hidde Ploegh**.

Submit a biomedical-research paper to *Nature* or other high-profile journals, and a common recommendation often comes back from referees: perform additional experiments. Although such extra work can provide important support for the results being presented, all too frequently it represents instead an entirely new phase of the project, or does not extend the reach of what is reported. It is often expensive and unnecessary, and slows the pace of research to a crawl. Among scientists in my field, there is growing concern that escalating demands by reviewers for the top journals, combined with the increasingly managerial role assigned to editors, now represents a serious flaw in the process of peer review.

Here, I offer some suggestions. The generalizations that follow have their pleasant exceptions, but the trend is that useful interventions are becoming exactly that — exceptions.

Rather than reviewing what is in front of them, referees often design and demand experiments for what would be better addressed in a follow-up paper. It is also commonplace for reviewers to suggest tests that, even if concluded successfully, do not materially affect conclusions. These are known in the trade as reviewer experiments. The demands seem to increase with the impact factor of a journal, as if referees feel that they need to raise the bar on the journal's behalf.

This has a serious and pernicious impact on the careers of young scientists, because it is not unusual for a year to pass before a paper is accepted into a high-profile journal. As a result, PhD degrees are delayed, postdocs may have to wait an entire year to compete for jobs and assistant professors can miss out on promotions.

The system also adds to tension between established, tenured lab heads charged with proper allocation of limited resources, and students and postdocs whose careers rely on papers in high-impact journals. The two sides will disagree on whether to cut their losses and consider lower-ranked journals, or to cave in to reviewers' demands.

The extra months of experiments increase costs for labs, without any obvious advantage for science. Although journals profit handily when prospective authors offer the best science possible, most do not spend money to produce it. For the publishing industry, this is an accepted business model, but it should come with greater responsibilities.

The scientific community should rethink how manuscripts are reviewed. Referees should be instructed to assess the work in front of them, not what they think should be the next phase of the project. They should provide unimpeachable arguments that, where appropriate, demonstrate the study's lack of novelty or probable impact, or that lay bare flawed logic

EXTRA MONTHS OF EXPERIMENTS INCREASE COSTS WITHOUT ANY OBVIOUS ADVANTAGE.

or unwarranted conclusions. They should abandon the attitude that screams: "look, I've read it, I can be as critical as the next dude and ask for something that's not yet in the manuscript", a reflexive approach to reviewing that has unfortunately become more or less standard. Many reviewers are also, of course, authors, who will receive such unreasonable demands in their turn, so why does the practice persist? Perhaps there is a sense of 'what goes around comes around', and scientists relish the chance to inflict their experiences on others.

The problem is made more acute by the unwillingness of editors to express their opinions. Instead, they consult an increasing number of reviewers (four or five is no longer an exception) in search of a majority opinion. Rather than taking a hard look at reviews and the experiments requested by referees, editors seem to default to the position that almost every requested experiment or revision can be justified. Editors often do not (or cannot?) assess revised manuscripts, and so send them out to reviewers again, losing more time and often bringing still more demands for further experiments.

I see three steps that journals can take to improve this deteriorating situation. First, they should insist that reviewers provide a rough estimate of the anticipated extra cost (in real currency) and effort associated with experiments they request. This is not unlike what all researchers are typically asked to provide in grant applications. Second, journals should get academic editors with expertise in the subject to take a hard look at whether the requests of reviewers will affect the authors' conclusions, and whether they can be implemented without undue delay. Third, reviewers should give a simple yes or no vote on the manuscript under scrutiny, barring fatal shortcomings in logic or execution. Once editors have decided that, in principle, the results are of interest to their publication and its readership (which is their editorial prerogative), passing a simple test of logical rigour and quality of data should be enough to get them through peer review. Multiple revisions rarely affect the overall conclusions of a study, as many an editor (and author, for that matter) would agree.

These changes would save time, speed exciting science to the public eye and provide much-needed clarity to authors — with significant savings to boot. Having read some of the biographies of the founders of molecular biology, it is hard to escape the impression that, once, the mechanics of science were indeed thus. It is worth revisiting the experiment, I should think. ∎

Hidde Ploegh *is a professor of biology at the Whitehead Institute of the Massachusetts Institute of Technology in Cambridge.*
e-mail: ploegh@wi.mit.edu

A. MAHON

ORGANIC CHEMISTRY

Alkanes to amines by bromine

Chemists have been seeking simple ways of inserting nitrogen or other elements into alkanes — simple hydrocarbons with just single bonds — to turn them into more useful molecules, such as drugs. Now Masahito Ochiai at the University of Tokushima in Japan and his colleagues show that a reactive bromine-based compound can be used for the addition of nitrogen without the traditional need for a metal catalyst or high temperature.

They used N-triflylimino-λ^3-bromane to insert nitrogen into a variety of alkanes at room temperature. Not only did this reaction achieve reasonable yields, in many cases it also selectively added nitrogen in just one of two possible positions.
Science 332, **448–451 (2011)**

APPLIED PHYSICS

Better X-ray vision

A new technique allows fainter features to be imaged by X-rays.

Conventional X-ray imaging relies on the absorption and scattering of X-ray photons by the object being imaged. But X-ray phase-contrast imaging instead detects changes in the photons' direction and velocity.

Alessandro Olivo and his colleagues at University College London used a conventional X-ray source outfitted with grated masks — one in front of the object for imaging and one behind it. The masks were offset slightly from one another so that they filtered out some of the photons, reducing background noise. The detector measures by how much photons have deviated from their path, capturing different image data from conventional X-ray imaging and boosting the visibility of fine detail.

The team used its technique to image biological specimens such as a beetle (**pictured**), as well as samples of interest for medical imaging, materials science and security inspection.
Appl. Optics 50, **1765–1769 (2011)**

C.A. SCHWEMM

ECOLOGY

More rain, more virus

Precipitation levels and species biodiversity may affect humans' exposure to zoonotic diseases — those carried by other animals.

Hantavirus, which is carried by wild rodents, especially deer mice (*Peromyscus maniculatus*, **pictured**), causes severe lung disease in humans. The prevalence of the Sin Nombre variant of hantavirus, which is endemic in deer mice on California's Channel Islands, is affected by several ecological factors.

John Orrock at the University of Wisconsin, Madison, and his colleagues analysed data from the islands and found that increased viral prevalence was linked to greater precipitation and island area, and also to fewer predator species. Precipitation accounted for 79% of the variation in prevalence. Adding in island area upped this to 93%, and including predator richness took the total to 98%.
Am. Nat. **doi:10.1086/659632 (2011)**

VIROLOGY

The keys to hepatitis entry

The identification of two proteins that grant the hepatitis C virus (HCV) access to host cells may provide new targets for potential drugs. At present, treatments for the virus, which causes liver disease, are limited.

Thomas Baumert at the University of Strasbourg in France and his team screened human liver cells for proteins that regulate HCV entry. They focused on two cell-surface proteins, EGFR and EphA2, which are key players in a regulatory network linked to HCV entry and are abundant in liver cells. Blocking the two proteins with inhibitors reduced HCV infection of cells in culture. In mice with HCV and transplanted human liver cells, an EGFR inhibitor slowed the rate of infection.
Nature Med. **doi:10.1038/nm.2341 (2011)**

PHARMACOLOGY

Antibiotics aided by other drugs

Combining antibiotics with drugs that do not directly kill bacteria may be one option in fighting antibiotic resistance.

The antibiotic minocycline blocks bacterial protein synthesis, but resistance has emerged in several human pathogens. Eric Brown and Gerard Wright of McMaster

University in Hamilton, Canada, and their group tested minocycline in combination with each of 1,057 approved drugs. They found 69 non-antibiotic compounds that boosted minocycline's ability to slow the growth of three bacterial species. Three of these drugs rendered multidrug-resistant clinical strains of the pathogen *Pseudomonas aeruginosa* susceptible to the antibiotic.

One of the three, the anti-diarrhoeal drug loperamide, may improve the uptake of minocycline and related antibiotics by the bacteria. The two combined reduced the bacterial load in mice infected with minocycline-resistant *Salmonella enterica* Typhimurium.
Nature Chem. Biol. doi:10.1038/nchembio.559 (2011)

GEOENGINEERING
Aerosols pose climate dangers

The risks of injecting aerosols into the stratosphere to combat global warming have been underestimated, say Klaus Keller and his team at Pennsylvania State University in University Park.

Some have proposed that adding fine particles and liquid droplets, which reflect heat, to the atmosphere could postpone the need for deep cuts to carbon emissions. But by modelling the economic effects of substituting aerosol injections for carbon cuts, the researchers show that potentially damaging side-effects of the aerosols — such as depletion of polar ozone and changes in precipitation patterns — could easily wipe out any benefit. And because aerosols disappear quickly whereas greenhouse gases linger, Earth could face even more abrupt and costly climatic change if any aerosol injections were interrupted by, for example, war or the breakdown of international pacts, the researchers say.
Clim. Change doi:10.1007/s10584-010-9961-z (2011)

BIOENGINEERING
Patching up the heart

Small patches of a protein matrix loaded with stem cells could help to repair the damage caused by a heart attack.

Gordana Vunjak-Novakovic and her colleagues at Columbia University in New York stripped human heart muscle tissue of its cells, leaving behind a porous protein scaffold. They filled this with human mesenchymal progenitor cells, which develop into muscle and other tissues, and then implanted patches of it into the damaged hearts of rat models of heart attack. The patches enhanced the growth of new blood vessels and stopped the left ventricle from enlarging and weakening — common effects of a heart attack.

Instead of forming new vessels or tissues, progenitor cells homed in on the damaged areas and released proteins that stimulated the native tissue to repair itself. Similar patches could deliver other types of cell, such as heart muscle cells created from reprogrammed stem cells, the researchers say.
Proc. Natl Acad. Sci. USA doi:10.1073/pnas.1104619108 (2011)

CLIMATE CHANGE
Sugar cane cools climate

The rapid expansion of sugar cane cultivation (**pictured**) for biofuels in Brazil may be cooling the local climate.

Scott Loarie at the Carnegie Institution for Science in Stanford, California, and his team analysed remote-sensing data for factors including land cover, temperature and albedo (reflectivity) in the Brazilian savannah, where sugar cane acreage is on the rise. They found that temperatures in the region increase by an average of 1.55 °C when natural vegetation is converted to non-cane crops or pasture, largely because of decreased water evaporation and transpiration (evaporation from plants). However, when planted on land already used for agriculture, sugar cane crops cool temperatures by 0.93 °C, thanks to greater evaporation and transpiration, and higher albedo.

The authors cite this as evidence that sugar cane's climate benefits will be greatest if restricted to existing agricultural lands and if the plant's cultivation does not drive deforestation elsewhere.
Nature Clim. Change doi:10.1038/nclimate1067 (2011)

P. FRIDMAN/CORBIS

ECOLOGY
Plant networks crowd neighbours

Plants that connect and spread through expansive underground networks reduce biodiversity, even in soils that are thought to boost it. This could mean that such 'clonal' species — some grasses, for example — will exacerbate plant biodiversity loss in soils that have been enriched by humans.

Fertilizing soil is known to lower plant-species richness, whereas natural variations in local soil nutrient levels are thought to increase it. Alexander Eilts at the University of Minnesota in St. Paul and his colleagues manipulated nutrient levels in outdoor plots for six years. When nutrient levels varied on a small scale, clonal species dominated by reaching between areas to get more nutrients. When nutrient levels varied over larger scales, or when clonal species were absent, biodiversity was higher.

Because of their underground networks, clonal plants may lower the ability of soil nutrient heterogeneity to support a diversity of plant species.
Am. Nat. doi:10.1086/659633 (2011)

◯ NATURE.COM
For the latest research published by *Nature* visit:
www.nature.com/latestresearch

The rapid progression of breast cancer has been linked to immune cells called macrophages. Now researchers have identified a protein that these 'tumour-associated' macrophages (TAMs) secrete that drives metastasis.

Erwei Song at Sun-Yat-Sen University in Guangzhou, China, Xuebiao Yao at the University of Science and Technology of China in Hefei and their colleagues found that the most abundant cytokine — a cell-signalling protein — produced by TAMs isolated from patients with invasive breast cancer is CCL18. Injection of this molecule into tumour-bearing mice boosted metastasis. Meanwhile, human breast cancer cells cultured with TAMs displayed a six- to ninefold increase in invasiveness. Adding an anti-CCL18 antibody tempered this response.
Cancer Cell 19, 541–555 (2011)

SEVEN DAYS *The news in brief*

European budget

Sticking to pre-agreed schedules for spending €50.5 billion (US$74 billion) on research over 2007–13, the European Commission has proposed a 13.3% rise in Europe's research budget next year, which would take it to €7.6 billion. The rise is included in the draft European Union budget for 2012, which was published on 20 April. The budget's suggested spending boost (of 4.9% from 2011) has struck some finance ministers as clashing with member states' economic austerity measures. But budget commissioner Janusz Lewandowski said that it was the "bare minimum" required to honour existing contracts. See go.nature.com/y4rd7u for more.

Altering embryos

Assisted-conception techniques that transfer genetic material between eggs to produce embryos free of inherited mitochondrial diseases seem, on the evidence so far, to be safe. That's the conclusion of a scientific report from the Human Fertilisation and Embryology Authority in London, published last week. The report calls for additional cell and animal experiments but does not consider the ethical and legal implications of such transfers, which are not yet permitted. They would create embryos with DNA from three sources — the mother, father and a mitochondrial DNA donor. See go.nature.com/yypt8o for more.

Italy's nuclear hold

Italy has halted plans to build four new nuclear reactors. They would have been the first to be built in the country since a referendum in 1987 shut down its nuclear power industry in the wake

S. SHOSTAK/SPL

Allen's alien hunt is halted

The Allen Telescope Array, which listens for signals of extraterrestrial life, has been shut down owing to a lack of funding. The University of California, Berkeley, has been operating the array of 42 radio dishes in northern California since 2007. But its funding from the National Science Foundation and the state of California has dropped, so the array has been put into hibernation. The SETI Institute in Mountain View, California, which co-runs the array, informed donors of the problem on 22 April, and said that it would prepare a fund-raising campaign. See go.nature.com/1dqfil for more.

of the Chernobyl disaster the previous year. After the accident at Japan's Fukushima nuclear plant in March, the Italian government introduced a one-year moratorium on the building plans. An amendment approved by the country's Senate on 20 April now extends this moratorium indefinitely.

Scientific integrity

The White House Office of Science and Technology Policy (OSTP) last week reported mixed progress for the scientific-integrity initiative that began with a memo issued by US President Barack Obama in 2009. The office said that 30 government agencies had responded to a December 2010 memo from John Holdren, director of the OSTP, in which he asked officials to prioritize the development of

integrity guidelines and report back in the next 120 days. But only six agencies have actually delivered draft policies, and the Holdren memo was itself delayed by 17 months relative to a timetable laid out by the president. See go.nature.com/eddwgh for more.

Mekong dam delay

The future of a controversial hydroelectric dam on the Mekong River in Laos remains unclear after the intergovernmental Mekong River Commission couldn't agree on whether to approve it. Laos wants to build the Xayaburi dam, which would be the first on the river's main channel outside China, and likely to be followed by a string of other dams if approved. But at a crunch decision meeting on 19 April, Cambodia,

Vietnam and Thailand raised doubts about the dam's ecological and social impacts. The decision has been deferred to an as-yet unscheduled meeting of the commission's council, which would involve environment ministers of the disputing nations.

Malaria thefts

The Global Fund to Fight AIDS, Tuberculosis and Malaria thinks that around US$2.5-million worth of its donated anti-malaria drugs have been stolen. The fund, based in Geneva, Switzerland, found the corruption among drug-procurement contracts totalling $98 million in 13 countries (mostly in Africa) for 2009–10. It is also checking contracts in earlier years.

Documents on the thefts were leaked last week to the Associated Press; the Global Fund had hoped not to publicize the investigation until it had been completed. The fund says that it has convened an international group of drug donors, which started meeting in February, to discuss how to stop thefts.

Oil-spill lawsuits

A year after the explosion of the Deepwater Horizon oil rig in the Gulf of Mexico, BP is suing three firms that were contracted to operate or make parts for the oil rig. Transocean of Geneva, Switzerland, which owned and operated the rig; Cameron International of Houston, Texas, which made the blowout preventer; and Halliburton of Houston, which designed the well's cement seal, will all be pursued for the roughly US$40 billion that BP has set aside for oil-spill costs. Companies had been given the deadline of 20 April — the one-year anniversary — to file lawsuits.

EVENTS

Fukushima zone

The Japanese government has started to enforce a 20-kilometre exclusion zone around the Fukushima nuclear plant, to stop people from returning to irradiated land. Roughly 27,000 households

are included in the zone. From 21 April, unauthorized entry was declared illegal; checkpoints block the routes in (**pictured**). At the Fukushima plant itself, robots are surveying the reactor cores while workers of the Tokyo Electric Power Company continue to inject cooling water into the reactors and to pump radioactive water out into storage buildings.

RESEARCH

Rhesus rewrite

Marc Hauser, an evolutionary psychologist at Harvard University in Cambridge, Massachusetts, has replicated findings from a 2007 paper that had been questioned because some raw data were missing. On 25 April the journal *Science* published the replication (J. N. Wood and M. D. Hauser *Science* doi:10.1126/science.1202596; 2011), which supports the original paper's conclusion that rhesus monkeys can discern the intentions underlying another individual's actions. Last

August, Harvard found Hauser guilty of eight counts of scientific misconduct; it emerged last week that he has been banned from teaching there for the coming year. See go.nature.com/yfifcj for more.

Collider record

The Large Hadron Collider — already the world's highest-energy particle accelerator — has set a new world record: for beam intensity. Also termed 'luminosity', this is a measure of the rate at which protons are colliding, and in turn determines how quickly collision data are generated. On 22 April, the beam luminosity at the collider, located at CERN, Europe's particle-physics laboratory near Geneva, Switzerland, reached 4.67×10^{32} collisions per square centimetre per second, passing the previous record of 4.024×10^{32}, held by the Tevatron collider at Fermilab in Batavia, Illinois.

Revamped ranking

The US National Research Council has revised its influential rankings of doctoral programmes in the United States after it discovered "four substantive errors" in the data. The much-studied rankings were originally released last September, and marked a major rethinking of a system that had last been updated in

COMING UP

29 APRIL
NASA's space shuttle *Endeavour* launches from the Kennedy Space Center in Florida for its final flight, which will be the penultimate mission of the shuttle fleet. It will carry the Alpha Magnetic Spectrometer, an experiment to measure cosmic rays and search for dark matter.
go.nature.com/1rsztj

30 APRIL−3 MAY
The American Physical Society meets in Anaheim, California. Key presentations include results from the Large Hadron Collider.
go.nature.com/uhzvez

1995 (see *Nature* **467,** 510; 2010). Most programmes' positions have not changed much in the revision — but in some fields, such as astrophysics and astronomy, around a third of programmes have seen their position ranges on various metrics change by more than 10%.

Oil-spill research

After months of delay, BP's fund of US$500 million for studies into the effects of last year's Gulf Coast oil spill has been opened up to researchers. On 25 April the Gulf of Mexico Research Initiative (GRI) announced that it was ready to review proposals from research consortia. BP announced the 10-year fund on 24 May 2010, but — apart from $40 million of fast-track funds — the GRI's roll-out of the cash had been hindered by bureaucracy and political manoeuvring (see *Nature* **470,** 317; 2011). Even now, grants will not be available until at least 30 August. See go.nature.com/6ki8rw for more.
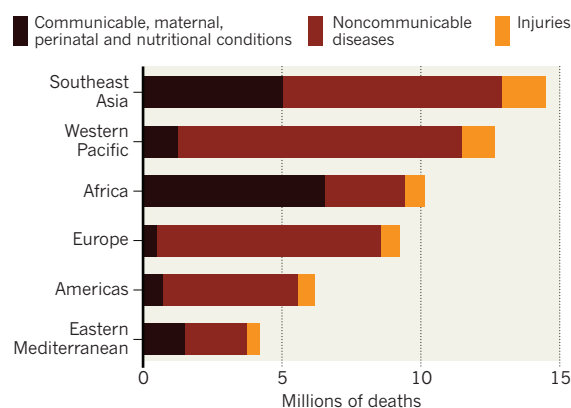
↻ **NATURE.COM**
For daily news updates see:
www.nature.com/news

TREND WATCH

The World Health Organization (WHO) says that levels of non-communicable diseases (NCDs) such as cardiovascular conditions, diabetes and cancers, have reached 'epidemic' proportions. In a 27 April report mapping the problem (see chart), it says that nearly 80% of NCD deaths occur in low- and middle-income countries, and such fatalities will rise by 15% globally by 2020. Many are caused by tobacco use, poor diet, physical inactivity or harmful use of alcohol, and are preventable, says the WHO.

THE CAUSES OF DEATH
Of the world's 57 million deaths in 2008, 63% were caused by non-communicable diseases (such as heart disease, diabetes and cancer).

Communicable, maternal, perinatal and nutritional conditions
Noncommunicable diseases
Injuries

Southeast Asia
Western Pacific
Africa
Europe
Americas
Eastern Mediterranean

0 — 5 — 10 — 15
Millions of deaths

▶ civilian nuclear plants, and offers assistance in times of crisis.

The WHO, in contrast, has an explicit remit to provide information during natural outbreaks of disease that spread across borders. This means that nations in a threatened region are eager for its expert assessments, says Duncan Snidal, a expert in international relations at the University of Oxford, UK. The IAEA, in comparison, actively probes the nuclear endeavours of members, including any undeclared weapons programmes, so they are reluctant to give it too much authority.

Nowhere has the IAEA's short leash been more apparent than in its International Nuclear and Radiological Event Scale. The agency introduced the numerical severity scale in 1990 to facilitate quick communication about nuclear accidents, but, bizarrely, it is national regulators that determine the rating of a particular emergency. Japan initially ranked the Fukushima crisis at level 5 by treating each reactor at the plant as a separate event, but a month after the event it grouped the reactors into a single incident and upped the rating to 7, the highest on the scale. The sudden change created confusion in the press and anxiety for the public, and showed just how inadequate the scale is, says Jeffrey Lewis, director of the East Asia nonproliferation programme at the James Martin Center for Nonproliferation Studies in Washington DC.

A NUCLEAR MINNOW
The IAEA has relatively limited resources compared with the United Nations agencies for food and health security.

BUDGET: US$ ● = 500 staff

IAEA International Atomic Energy Agency — $0.5 bn — 2,300
Of which safety and security: US$42 million

FAO Food and Agriculture Organization — $1 bn — 3,600

WHO World Health Organization — $4.5 bn — 9,000

Whether the IAEA can gain more independence to assess nuclear accidents is uncertain, but its emergency scale is very likely to be revisited given the confusing way in which it has been used. Some commentators have also called for a stronger IAEA that can enforce global safety standards or take control in an emergency, but Acton labels such ideas "insane". Nations might begin to feel that "safety is not our responsibility", he says. Moreover, a nuclear plant's operators are often better placed to handle a crisis than are outside officials unfamiliar with the facility. In any event, Acton doubts that countries will abdicate oversight of their reactors. "The whole thing is fanciful," he says.

Lewis says that a more outspoken IAEA could have dispelled some of the confusion around the impact of Fukushima. Japanese authorities played down the severity of the accident, for example, while US regulators on the scene described a much more frightening worst-case scenario. But if the IAEA is just one more voice in the tumult, "and if it is incompetent, then it makes everything much, much worse", Lewis adds.

It will ultimately be up to the member states to determine how much independence the IAEA will have in a future crisis. The agency's own position is unknown: it declined *Nature*'s interview request for this story. ■ **SEE EDITORIAL P.389**

ENERGY

Is FutureGen betting on the wrong rock?

Commercial rival says gas spreading and leakage could harm US carbon-sequestration effort.

BY JEFF TOLLEFSON

Kurt Zenz House watches from a corner office in Berkeley, California, as carbon dioxide is pumped into sandstone 2,000 metres below the southern Illinois farmland. A colourful plume grows on his computer screen, modelling the movement of CO_2 through the porous rock. As decades elapse in the simulation, the CO_2 rises and settles under the dome of shale that caps the formation.

House says that this is exactly what should happen. "CO_2 injected anywhere inside the geological boundary will stay within the boundary," he says. And, he argues, that is exactly what won't happen when FutureGen 2.0, the US government's main commercial demonstration project for carbon capture and sequestration (CCS), begins to store CO_2 from a coal-fired power plant in Illinois in 2015.

House's company, C12 Energy in Berkeley, is challenging the strategy that the project has chosen for sequestering the 1.3 million or so tonnes of CO_2 that it aims to capture each year. In its demonstration effort, FutureGen intends to inject the CO_2 into porous sandstone with no dome or other structure to help trap the gas — arguing in part that such formations are too rare to rely on for large-scale storage. But C12 and its backers, who advocated a rival site, say that the strategy risks legal and financial problems if the CO_2 spreads farther than expected.

Last February, FutureGen rejected C12's bid to handle the storage side of the

⟳ NATURE.COM
For more on the challenges of carbon sequestration see:
go.nature.com/ks9agi

project, so C12 is pushing ahead with plans to develop its own commercial sequestration operation. It is acquiring rights to pump CO_2 into its Illinois site — a domed reservoir in Fayette County 160 kilometres from the plant (see 'Tilted playing field') — and into reservoirs in eight other states in the hope that CCS will eventually become commonplace at other power plants.

Killed by US President George W. Bush in 2008, FutureGen was resurrected a year later by President Barack Obama as a centrepiece of his administration's strategy for climate change and energy. The project is a US$1.3-billion deal between the US Department of Energy and a group of major energy companies to capture and store CO_2 from the exhaust of a coal plant near Meredosia, Illinois. After rejecting C12's bid, FutureGen settled on a saline aquifer 1,200–1,500 metres deep in Morgan County, where the plant is located.

As well as lacking a structural trap, the sandstone in Morgan County sits at a slight angle and is shallower than that in Fayette County. All of this makes gas injected there likely to be more mobile and less predictable, says Daniel Schrag, a geochemist at Harvard University in Cambridge, Massachusetts. That could "poison the well" for CCS if legal problems ensue, says Schrag, who was House's PhD adviser and now chairs C12's board of science advisers.

By trapping the injected CO_2 with geological structures, C12 says that it can minimize the number of leases for subsurface rights it needs to negotiate with landowners. In its bid, submitted in partnership with local governments, the company had offered to charge a flat fee of $140 million to handle everything at the storage site. The pipeline to carry the CO_2 from the plant would have cost another $100 million to $200 million. FutureGen initially declined to talk about finances, saying only that the project would come in at or below budget. It now says that its approach will be cheaper, with a maximum budget of around $125 million for the sequestration work.

Kenneth Humphreys, chief executive of the FutureGen Alliance, says that the geology in Morgan County is both suitable and typical of what is readily available around the world. "We are focused on what could well be the workhorse solution," Humphreys says. C12's site is "interesting, but not nearly as common".

WIDE SPREAD

Outside experts say that it is hard to compare the sites until both have been fully characterized with modern seismic and other technologies. Although the type of structural containment proposed by C12 has advantages, it could be a disadvantage if a leak develops because all the CO_2 would escape from one place, says Susan Hovorka, a carbon-sequestration expert at the University of Texas at Austin.

Similarly, although a lack of containment might translate into a larger plume, the dispersed CO_2 is more likely to dissolve into groundwater or to remain trapped in pore space, adds Robert Finley, who heads the Advanced Energy Technology Initiative at the Illinois State Geological Survey. "The C12 guys have one approach," Finley says, "but I don't think the validity of their arguments necessarily negates other approaches."

But C12 and its supporters argue that structural closure lessens the risk of liability if CO_2 migrates into pore space for which the project has not secured legal access. "Let's start with a viable project where the risk is as low as possible," says Tony Meggs, a visiting engineer at the Massachusetts Institute of Technology in Cambridge, who chairs C12's board of directors.

Humphreys says that FutureGen plans to sign lease agreements for subsurface rights with the landowners who own the land above the plume's location,

> *"Let's start with a viable project where the risk is as low as possible."*

but that will be more difficult if the plume spreads. When initially considering potential sites, FutureGen assumed a plume size of roughly 4 square kilometres. Humphreys has since acknowledged that the CO_2 could end up spreading further, but the alliance won't know for sure until they have studied the geology of the site in detail. House says his team's modelling suggests that the gas could spread across 80–160 square kilometres in FutureGen's site.

House and his C12 colleagues are taking a broader approach, seeking legal access to all of the pore space in their reservoir. So far, C12 has signed leases with nearly 200 landowners, covering roughly 60% of the 100 square kilometres in their field, an area that could store roughly 13 times more CO_2 than FutureGen expects to capture. Meanwhile, the company is working with Illinois lawmakers on extending to carbon storage the laws that apply to oil and gas fields, which grant legal access to the entire area once a certain percentage of the owners involved have signed on.

Some say that C12's business model simply didn't fit into the FutureGen project. James Wood, deputy assistant secretary for clean coal programmes at the Department of Energy, says that handing the storage half of the FutureGen project over to a commercial entity such as C12 might undercut the mission of the FutureGen Alliance partners, who are bolstering the more than $1 billion in public money with roughly $244 million of their own. "We don't want anybody skipping out halfway through the project," Wood says.

But for Schrag, that's further evidence that FutureGen is not being run the way it should be, which is as a commercial venture. He fears that FutureGen's current course could inflate costs and ultimately set the industry back. "This is the government's flagship demonstration of carbon capture and sequestration," Schrag says. "We really need it to work." ∎
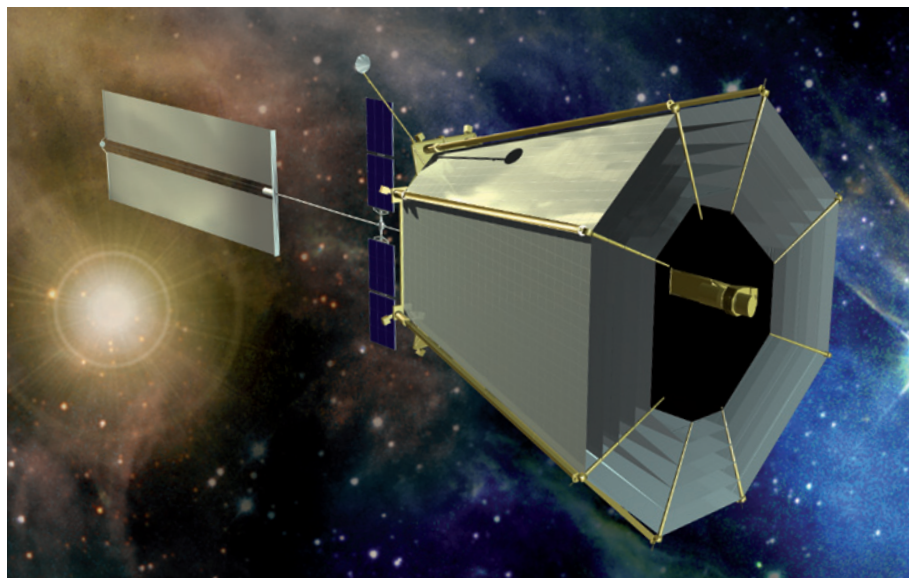
TILTED PLAYING FIELD

The sandstone formation is shallower under Morgan County, Illinois, where FutureGen plans to store its carbon, than in an alternative storage site in Fayette County.

Approximate depth in metres
- ≤600
- 600–1,200
- 1,200–1,800
- 1,800–2,400
- 2,400–3,000
- ≥3,000
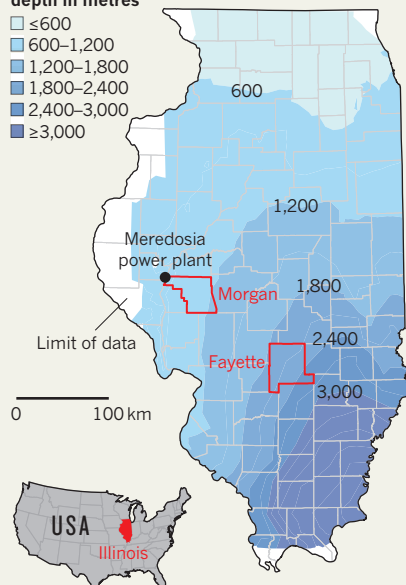
Exoplanet hunters want something to replace the postponed Terrestrial Planet Finder.

SPACE SCIENCE

Astronomers mull merger of missions

Cosmic–origins scientists convene with exoplanet hunters.

BY EUGENIE SAMUEL REICH

NASA's constrained budget is encouraging some creative pairings. This week, scientists eager to find other habitable worlds explored the possibility that a future space telescope for probing the origins of stars and galaxies could serve their needs as well.

For months, NASA's Cosmic Origins Program Analysis Group (COPAG) has been discussing the prospects for a large ultraviolet–visible-wavelength telescope that would fly sometime after 2020 and explore the architecture of galaxies and the details of their formation. Then the Exoplanet Exploration Program Analysis Group (ExoPAG) got wind of the plan. Astronomer Kenneth Sembach of the Space Telescope Science Institute in Baltimore,

Maryland, says that he was originally expecting a small group at a 26 April joint COPAG and ExoPAG meeting he is co-hosting. In the end, 49 astronomers attended in person, online or by dial-in. "It's taken on a life of its own," he says.

Exoplanet hunters are keen for a new instrument. Results are pouring in from Kepler, NASA's first orbiting telescope dedicated to finding other Earths. Although Kepler can detect planets between about 184 and 920 parsecs away that are transiting in front of their host stars, it is generally unable to determine their masses or compositions. Researchers hope for a follow-up mission that would image Earth-like planets around the nearest 100 Sun-like stars at sufficient resolution to detect traces of oxygen or water in their atmospheres.

The Terrestrial Planet Finder — a NASA

project along these lines — was indefinitely postponed in 2007, but the agency has continued to put US$6 million a year into developing technology for exoplanet searches. Now planet hunters think that joining forces with COPAG will be the winning strategy. Jim Kasting, a planetary scientist at Pennsylvania State University in University Park who is chair of the exoplanet group, says the two communities would both like to see a 4–8-metre telescope in space that would cost in excess of $5 billion. "Our interests are basically aligned," he says.

Such a mission would compete for top billing in the next decadal survey of astronomy by the US National Academy of Sciences, due in 2020. The big question, which follow-up meetings will consider, is whether the same technology can do both kinds of science.

A cosmic-origins mission would need to collect as much ultraviolet and visible light as possible to image intergalactic gas, star formation and Sun-like stars in nearby galaxies. A planet-hunting probe would need a coronagraph to block direct light from host stars, and would have to be sensitive to the visible and near-infrared wavelengths that Earth-like planets primarily emit. The infrared 6.5-metre James Webb Space Telescope (JWST), scheduled for launch in 2014, will be able to see larger planets but will not be sensitive to Earth-sized ones. The key to making the joint concept work will be developing a reflective coating for the telescope's mirror that works from the ultraviolet to the infrared ranges and does not distort the incoming light waves in a way that would make it difficult to tease out the image of an exo-Earth.

Sara Seager, a planet hunter at the Massachusetts Institute of Technology in Cambridge, is not convinced by the strategy. "It sounds greedy, but you're cheating yourself if you make a mission everyone has to share," she says. Although smaller-scale planet-hunting missions set to launch in the next decade will not be sensitive enough to find other Earths, she adds, they may help increase momentum to build a dedicated space telescope that can.

But Natalie Batalha of San José State University in California, deputy team leader for Kepler, says a joint mission may be the logical response to NASA's money troubles, which cost overruns of the JWST have exacerbated. "These projects are very expensive and ambitious and we have to find ways to leverage off other programmes," she says. ∎

MORE ONLINE

STEM CELLS

California ponders cell–banking venture

State agency grapples with technical and ethical challenges.

BY ERIKA CHECK HAYDEN

Investigators creating a wealth of adult-derived stem-cell lines may soon get new banks for storing them. Unlike stem cells derived from human embryos, induced pluripotent stem (iPS) cells face few political and ethical hurdles because they can be produced from adult tissue. As a result, investigators have created a rash of iPS-cell lines representing a huge range of genetic variations.

Now, one of the biggest funders of stem-cell research, the California Institute for Regenerative Medicine (CIRM) in San Francisco, is contemplating ways to organize these efforts. The agency is holding a trio of meetings this week to consider funding iPS-cell banking projects. This is a welcome development for scientists hoping to use the cell lines to test responses to new drugs and, eventually, to grow replacement tissues for treating degenerative diseases.

"We're creating these lines without really having a process to distribute them, and without really knowing whether people want to use them," says Jeanne Loring, an iPS-cell researcher at the Scripps Research Institute in La Jolla, California. "For the most part, iPS-cell banking efforts are fragmentary."

CIRM is not the only organization trying to change that. On 20 April, the public–private Innovative Medicines Initiative in Brussels announced that it is considering funding an iPS-cell bank. And in Japan, Shinya Yamanaka — the researcher at Kyoto University who ignited the field with his breakthrough creation of iPS cells from mice in 2006 — is developing a bank using the premise that cells taken from 50 carefully selected donors would be immunologically matched to 90% of the Japanese population.

In the United States, some banking efforts are already under way, such as one funded by the state of Massachusetts; some institutions, such as the Harvard Stem Cell Institute in Cambridge, also distribute lines created by their investigators. CIRM may build on an existing effort by the US National Institute of Neurological Disorders and Stroke (NINDS) in Bethesda, Maryland, which is banking iPS cells derived from people with three neurodegenerative diseases at the Coriell Institute for Medical Research in Camden, New Jersey.

But to succeed, CIRM and other would-be bankers must tackle some important issues. For a start, researchers use many different methods to generate iPS cells. A potential cell bank would need to decide how to handle such heterogeneity, because it affects the potential uses of a cell line.

The cell lines themselves also vary widely, especially in their pluripotency — their ability to develop into a range of tissues. And because studies have shown that iPS cells can acquire mutations during the reprogramming process and while they are in culture, an iPS-cell bank would need to monitor them constantly using genomic tests.

NINDS has tackled some of these issues through its banking project, which is depositing 25 iPS-cell lines in a small repository at the Coriell Institute. Coriell carries out a battery of tests on cell lines submitted to the bank, such as chromosome typing to monitor the potential accumulation of mutations. It is also subjecting each line to a set of genomic and epigenomic tests for pluripotency, which

> *"We're creating these lines without really having a process to distribute them."*

have been developed by Kevin Eggan, a stem-cell researcher at Harvard University in Cambridge, Massachusetts. CIRM looks set to emulate this approach by using a 'scorecard' of assays to characterize each line in its bank.

Perhaps more complicated than the scientific issues are the policy issues, such as what level of consent should be required from individuals who donate cells. Such issues are "on the same order of complexity as the issues surrounding derivation and use of human embryonic stem cells", says stem-cell researcher Roger Pedersen of the University of Cambridge, UK. Although iPS cells do not require the destruction of an embryo, the work that is done with them could yield private genetic information or great commercial benefit — both of which are potential concerns for the living adult from whom the cell line is derived.

CIRM has dedicated one of the meetings this week to these ethical issues. Given the scale of the questions — and the hope that CIRM's efforts will smooth the way for other cell banks — others in the field will be watching closely. ∎

CURSE OF THE PHARAOH'S DNA

Some researchers claim to have analysed DNA from Egyptian mummies. Others say that's impossible. Could new sequencing methods bridge the divide?

BY JO MARCHANT

ameras roll as ancient-DNA experts Carsten Pusch and Albert Zink scrutinize a row of coloured peaks on their computer screen. There is a dramatic pause. "My god!" whispers Pusch, the words muffled by his surgical mask. Then the two hug and shake hands, accompanied by the laughter and applause of their Egyptian colleagues. They have every right to be pleased with themselves. After months of painstaking work, they have finally completed their analysis of 3,300-year-old DNA from the mummy of King Tutankhamun.

Featured in the Discovery Channel documentary *King Tut Unwrapped* last year and published in the *Journal of the American Medical Association* (*JAMA*)[1], their analysis — of Tutankhamun and ten of his relatives — was the latest in a string of studies reporting the analysis of DNA from ancient Egyptian mummies. Apparently revealing the mummies' family relationships as well as their afflictions, such as tuberculosis and malaria, the work seems to be providing unprecedented insight into the lives and health of ancient Egyptians and is ushering in a new era of 'molecular Egyptology'. Except that half of the researchers in the field challenge every word of it.

Enter the world of ancient Egyptian DNA and you are asked to choose between two alternate realities: one in which DNA analysis is routine, and the other in which it is impossible. "The ancient-DNA field is split absolutely in half," says Tom Gilbert, who heads two research groups at the Center for GeoGenetics in Copenhagen, one of the world's foremost ancient-DNA labs.

Mummies found in King Tutankhamun's tomb are at the centre of a dispute over DNA analysis.

Unable to resolve their differences, the two sides publish in different journals, attend different conferences and refer to each other as 'believers' and 'sceptics' — when, that is, they're not simply ignoring each other. The Tutankhamun study reignited long-standing tensions between the two camps, with sceptics claiming that in this study, as in most others, the results can be explained by contamination. Next-generation sequencing techniques, however, may soon be able to resolve the split once and for all by making it easier to sequence ancient, degraded DNA. But for now, Zink says, "It's like a religious thing. If our papers are reviewed by one of the other groups, you get revisions like 'I don't believe it's possible'. It's hard to argue with that."

RISE AND FALL

The disagreement stems from the dawn of ancient-DNA research. In the 1980s, a young PhD student called Svante Pääbo worked behind his supervisor's back at the University of Uppsala in Sweden to claim he had done what no one else had thought was possible: clone nuclear DNA from a 2,400-year-old Egyptian mummy[2]. Soon researchers realized that they could use a new technique called polymerase chain reaction (PCR) to amplify tiny amounts of DNA from ancient samples. There was a burst of excitement as DNA was reported from a range of ancient sources, including insects preserved in amber and even an 80 million-year-old dinosaur[3].

Then came the fall. It turned out that PCR, susceptible to contamination at the best of times, is particularly risky when working with tiny amounts of old, broken-up DNA. Just a trace of modern DNA — say from an archaeologist who had handled a sample — could scupper a result. The 'dinosaur' DNA belonged to a modern human, as did Pääbo's

> "I don't understand people's harshness. This is pioneering work."

pioneering clone. Once researchers began to adopt rigorous precautions[4], including replicating results in independent labs, attempts to retrieve DNA from Egyptian mummies met with little success[5].

That's no surprise, say sceptics. DNA breaks up over time, at a rate that increases with temperature. After thousands of years in Egypt's hot climate, they say, mummies are extremely unlikely to contain DNA fragments large enough to be amplified by PCR. "Preservation in most Egyptian mummies is clearly bad," says Pääbo, now at the Max Planck Institute for Evolutionary Anthropology in Leipzig and a leader in the field. Ancient-DNA researcher Franco Rollo of the University of Camerino in Italy went so far as to test how long mummy DNA might survive. He checked a series of papyrus fragments of various ages, preserved in the similar conditions to the mummies. He estimated that DNA fragments large enough to be identified by PCR — around 90 base pairs long — would have vanished after only around 600 years[6].

Yet all the while, rival researchers have published a steady stream of papers on DNA extracted from Egyptian mummies up to 5,000 years old. Zink and his colleagues have tested hundreds of mummies, and claim to have detected DNA from a range of bacteria, including *Mycobacterium tuberculosis*, *Corynebacterium diphtheriae* and *Escherichia coli*, as well as the parasites responsible for malaria and leishmaniasis.

In a high-profile study last year, a team led by microbiologist Helen Donoghue at University College London reported finding DNA from *M. tuberculosis* in Dr Granville's mummy[7] — named after physician Augustus Granville, the first person to autopsy a mummy, in 1825.

In the case of tuberculosis (TB) at least, Donoghue vehemently disagrees with the idea that DNA can't survive in Egyptian mummies. Mycobacteria such as *M. tuberculosis* have cell walls that are rich in lipids, which degrade slowly and protect the DNA, she argues. Donoghue claims that in many cases she has confirmed the presence of the bacterium by detecting these lipids directly. She says the extreme anti-contamination measures demanded by the big ancient-DNA labs are not as vital for ancient microbial DNA as they are for human DNA. After all, she says, modern diagnostic labs routinely detect TB using PCR — which suggests that the test is not as susceptible to contamination as the sceptics fear. In Donoghue's view, "some of the precautions

they talk about are totally over the top compared to every diagnostic lab in the country".

The sceptics are unmoved. Without highly stringent controls in place, it's impossible to show that any microbial sequences are from ancient DNA and not from related modern microbes, says Gilbert. "How do you know you've got TB and not some other bacterium with a similar DNA sequence?" He and other critics believe that this entire body of research is based on wishful thinking.

The two groups have now grown tired of arguing. "It's largely dealt with by ignoring each other," says Ian Barnes, a molecular palaeontologist at Royal Holloway, University of London, who works on DNA from ancient animals, including mammoths. "There's enough dead stuff around, you're not obliged to get into anyone else's area."

A ROYAL ARGUMENT

After the *JAMA* study on Tutankhamun and his family, however, the arguments resumed in force. Studies of human DNA from Egyptian mummies are the most controversial of all. One reason is the high profile of the claims. Another is that contamination from modern human DNA is excruciatingly difficult to detect, because its genetic make-up is almost identical to that of a human mummy's. On top of that, restricted access to samples makes it hard to check any claims in an independent lab. After more than a century in which valuable artefacts flooded out of the country to museums and private collections all over the world, the Egyptian authorities imposed a ban on removing archaeological samples from Egypt. Most non-Egyptian researchers wanting to study mummies are limited to museum exhibits elsewhere.

The Tutankhamun project was carried out by an Egyptian team recruited by archaeologist Zahi Hawass, Egypt's top official in charge of antiquities. It was the first ancient-DNA study on royal mummies, and the country lacked the necessary expertise. So Hawass asked Zink, a prominent researcher at the EURAC Institute for Mummies and the Iceman in Bolzano, Italy, and Pusch, of the University of Tübingen, Germany, to act as consultants. The pair designed and oversaw the study, including the building of two dedicated labs in Cairo. The labs were partly paid for by the Discovery Channel, which filmed the project.

The researchers deny that the television involvement put them under excessive pressure to produce dramatic results. But working for the cameras did make a challenging project even tougher, says Pusch. "Each time they came in to film, we had to close the lab for a week to clean." Eventually the TV crew was banished and the lab scenes reconstructed.

In the end, the project seemed to be a wild success, and its findings drew wide press attention. The researchers claimed to have detected DNA from the malaria parasite *Plasmodium falciparum* in several of the mummies, including Tutankhamun, suggesting that the infection had contributed to their deaths. They also said they had retrieved

Archaeologist Zahi Hawass with King Tut's grandmother and a lot of press.

fragments of human DNA from every mummy tested and used the data to construct a five-generation family tree, from Tutankhamun's great-grandparents to the two tiny bodies found in his tomb, identified as his stillborn children.

The whole episode has only raised eyebrows in the other half of the community. "I'm very sceptical," says Eske Willerslev, director of Copenhagen's Center for GeoGenetics, who co-authored a letter to *JAMA* disputing the results[8]. His major concern, shared by others, was the method of DNA analysis used. Rather than extracting and sequencing DNA, the team used a technique called genetic fingerprinting, which involves measuring the size of the DNA products that have been amplified by PCR. It is rarely used in ancient-DNA studies, say critics, because without sequence data it is especially difficult to rule out contamination. And on a well-handled mummy such as Tutankhamun, say sceptics, contamination could be rife.

BONES OF CONTENTION

The Tutankhamun team carried out many controls, including replication of the tests by different teams in the two labs and comparing the mummy DNA fingerprints with those of the research team to cross-check for contamination. Zink and Pusch add that the samples were taken from within the mummies' bones where, they say, contaminating DNA should not have reached.

Zink and Pusch think that the mummification process protected the DNA from degrading in the hot tomb by removing water, which is required for the main mechanism of DNA decay, called depurination. Egyptian embalmers dried bodies with natron, a naturally occurring mixture of salts, immediately after death. "The Egyptians really knew how to preserve a body," says Zink. "They got rid of the water very fast." Tutankhamun was also smothered with embalming and anointing materials, thought to contain ingredients such as bitumen, plant oils and beeswax, and Pusch believes it gave the DNA additional protection from the damaging effects of water. Hawass was not directly involved in the DNA research, but he stands by the team's conclusions, saying that the DNA in Egyptian mummies seems to be well preserved.

"There are a number of things right about the paper," says David Lambert, an ancient-DNA researcher and evolutionary biologist at Griffith University in Nathan, Queensland. Lambert points out that the Tutankhamun team was not able to amplify Y-chromosome markers from the female mummies, which argues against contamination from modern archaeologists, who are generally male. In unpublished work, he says he has amplified DNA from mummified ibises, a sacred bird in ancient Egypt. "We're confident that traditional PCR methods work with some of the material that we've got," he says.

Sceptics, however, doubt that there was sufficient DNA left in Tutankhamun for the result to be real. They say that a mummified body would soon soak up any moisture available in the atmosphere, especially into its porous bones. When British archaeologist Howard Carter first opened Tutankhamun's coffins in 1925, he reported that they had been damaged by humidity. But it is difficult for anyone else to replicate the DNA work without permission to access the samples.

The Tutankhamun study has left the field more divided than ever, with clear frustration on both sides. "I don't understand people's harshness," Pusch says. "This is pioneering work." He and Zink say that they are sequencing DNA from the mitochondria and Y chromosomes of the mummies, and plan to publish these results this year.

But now, after years of conflict, strides in sequencing technology are changing the game. The newest techniques can read much shorter fragments — easily down to the 30 base pairs that might be found in a 2,000-year-old Egyptian mummy. "That pushes the [DNA] survival time a long way back," says Gilbert. "Things that we wrote off in the past, we can now get genomes on." And, crucially, the speed of the techniques makes it much easier to sequence a sample multiple times and to rule out contamination by checking for patterns of damage characteristic of ancient DNA.

↪ **NATURE.COM**
Listen to a podcast about the mummy DNA dispute at:
go.nature.com/x44ila

Tiny bodies buried with Tutankhamun are thought to be his stillborn children.

Last year, these techniques enabled Willerslev, Gilbert and their colleagues to publish the full genome sequence of a palaeo-Eskimo from Greenland that is some 4,000 years old[9]. Within weeks, teams led by Pääbo published the genome of a 38,000-year-old Neanderthal[10] and a previously unknown hominin from southern Siberia[11]. Meanwhile Zink's team is on the brink of publishing the genome of Ötzi the Iceman.

All these specimens were preserved in the cold — but Willerslev is already using next-generation techniques to extract DNA from various South American mummies, some of which have been preserved in warmer conditions. "Some are definitely working," he says. But, he adds, he is finding tremendous variability in whether samples yield DNA — a possible reason why Egyptian mummies have yielded such conflicting results. With the cost of sequencing falling sharply, researchers are lining up to try the techniques on Egyptian mummies.

Zink and Pusch are now negotiating the complex political path towards using next-generation techniques on Tutankhamun and his kin. "We would love to do this," says Zink. "It would absolutely make sense. The problem is to do it in Egypt." With no samples allowed out of the country, they would have to take the sequencing machines to Cairo, an expensive proposition. And there is concern, says Zink, that such work might yield politically sensitive information about the genetic origin of the pharaohs, and whether any of their descendants are alive today. "This goes right to their history."

Still, Zink is optimistic that next-generation sequencing will help to bring the fractured field back together. "I think it is really time to bring together the different sides and stop arguing about each other's work," he says. "With next-generation sequencing, people can't just say 'I don't like it'. People have to discuss the work based on the data themselves." Willerslev agrees, offering a rare olive branch. "I think we will find that the believers have been too uncritical," he says. "But the sceptics have probably been too conservative." ∎

Jo Marchant *is author of* Decoding the Heavens: Solving the Mystery of the World's First Computer.

1. Hawass, Z. *et al. J. Am. Med. Assoc.* **303,** 638–647 (2010).
2. Pääbo, S. *Nature* **314,** 644–645 (1985).
3. Woodward, S. R., Weyand, N. J. & Bunnell, M. *Science* **266,** 1229–1232 (1994).
4. Cooper, A. & Poinar, H. *Science* **289,** 1139 (2000).
5. Krings, M. *et al. Am. J. Hum. Genet.* **64,** 1166–1176 (1999).
6. Marota, I., Basile, C., Ubaldi, M. & Rollo, F. *Am. J. Phys. Anthropol.* **117,** 310–318 (2002).
7. Donoghue, H. D. *et al. Proc. R. Soc. B* **277,** 51–56 (2010).
8. Lorenzen, E. D. & Willerslev, E. *J. Am. Med. Assoc.* **303,** 2471 (2010).
9. Rasmussen, M. *et al. Nature* **463,** 757–762 (2010).
10. Green, R. E. *et al. Science* **328,** 710–722 (2010).
11. Reich, D. *et al. Nature* **468,** 1053–1060 (2010).

COMMENT

Emergency drills such as this mislead the public into believing that the Tokai district is due a magnitude-8 quake soon.

Shake–up time for Japanese seismology

Robert J. Geller calls on Japan to stop using flawed methods for long–term forecasts and to scrap its system for trying to predict the 'Tokai earthquake'.

For the past 20 years or so, some seismologists in Japan have warned of the seismic and tsunami hazards to the safety of nuclear power plants, most notably Katsuhiko Ishibashi, now professor emeritus at Kobe University. Their warnings went unheeded. Yet in the immediate aftermath of the magnitude-9.1 earthquake that struck Tohoku on 11 March, pundits could be found on many Japanese TV stations saying that it was "unforeseeable".

The 'foreseen' earthquakes were presumably the hypothetical future earthquakes used by the Japanese government to produce national seismic hazard maps for Japan[1]. The modellers assume that 'characteristic earthquakes' exist for various zones, choose the fault parameters for each zone as the input to their model, and then produce probabilistic hazard maps.

Although such maps may seem authoritative, a model is just a model until the ▶

SUMMARY

● The Japanese government should admit to the public that earthquakes cannot be reliably predicted.
● Use of the misleading term 'Tokai earthquake' should cease.
● The 1978 Large-Scale Earthquake Countermeasures Act should be repealed.

KYODO/NEWSCOM

methods used to produce it have been verified. The regions assessed as most dangerous are the zones of three hypothetical 'scenario earthquakes' (Tokai, Tonankai and Nankai; see map). However, since 1979, earthquakes that caused 10 or more fatalities in Japan actually occurred in places assigned a relatively low probability. This discrepancy — the latest in a string of negative results for the characteristic earthquake model and its cousin, the seismic-gap model[2–4] — strongly suggests that the hazard map and the methods used to produce it are flawed and should be discarded.

Globally, in the past 100 years, there have been five subduction-zone earthquakes of magnitude 9 or greater (Kamchatka 1952, Chile 1960, Alaska 1964, Sumatra 2004, Tohoku 2011), which suggests that the upper limit on the possible size of a subduction-zone earthquake may not much depend on the details of the subduction modality[5]. Large tsunamis have frequently struck the Pacific coast of the Tohoku district. The well-documented 1896 Sanriku tsunami had a maximum height of 38 metres and caused more than 22,000 deaths. The 869 Jogan tsunami is documented to have had a height roughly comparable to, or perhaps slightly less than, that of the 11 March tsunami.

If global seismicity and the historical record in Tohoku had been used as the basis for estimating seismic hazards, the 11 March Tohoku earthquake could easily have been 'foreseen' in a general way, although not of course its particular time, epicentre or magnitude. Countermeasures for dealing with such events could and should have been incorporated in the initial design of the Fukushima nuclear power plants.

THE 'TOKAI EARTHQUAKE'
In the 1960s, plate tectonics became generally accepted as the fundamental paradigm of solid-Earth geoscience. Researchers in several countries made efforts to combine plate tectonics with seismicity data to make long-term forecasts of large earthquakes. The idea was very simple. It was hypothesized that zones where no large earthquakes had occurred for a while, dubbed 'seismic gaps', were ripe for imminent large events. However, the seismic-gap hypothesis failed the test of reality[2]. Over tens of thousands of years or longer, the net slip released by earthquakes and aseismic slip must match net inter-plate motion. But we now know that this catching-up process does not occur regularly or cyclically, as is further underscored by the 11 March earthquake.

In the mid-1970s, when enthusiasm for the seismic-gap model was still widespread in the global geoscience community, several researchers in Japan proposed that the plate boundary off the Tokai district was a seismic gap where a magnitude-8 earthquake could be expected[6]. The neighbouring Tonankai and Nankai districts were also labelled as being seismic gaps[7]. No large earthquake has occurred in any of these districts since 1975, but they are still classified as the most hazardous regions in the country by the Japanese government (see map).

Over the past 30 years or so, government spokesmen and university scientists associated with the government's Headquarters for Earthquake Research Promotion (or its various predecessors) have used the term 'Tokai earthquake' so often that the public and news media have come to view it as a 'real earthquake' rather than merely an arbitrary scenario (1.78 million hits in a Japanese-language Google search). This misleads the public into believing that the clock is ticking down inexorably on a magnitude-8 earthquake that is certain to strike the Tokai district in the near future. Use of the term 'Tokai earthquake' (and its companions 'Tonankai earthquake' and 'Nankai earthquake') should therefore cease.
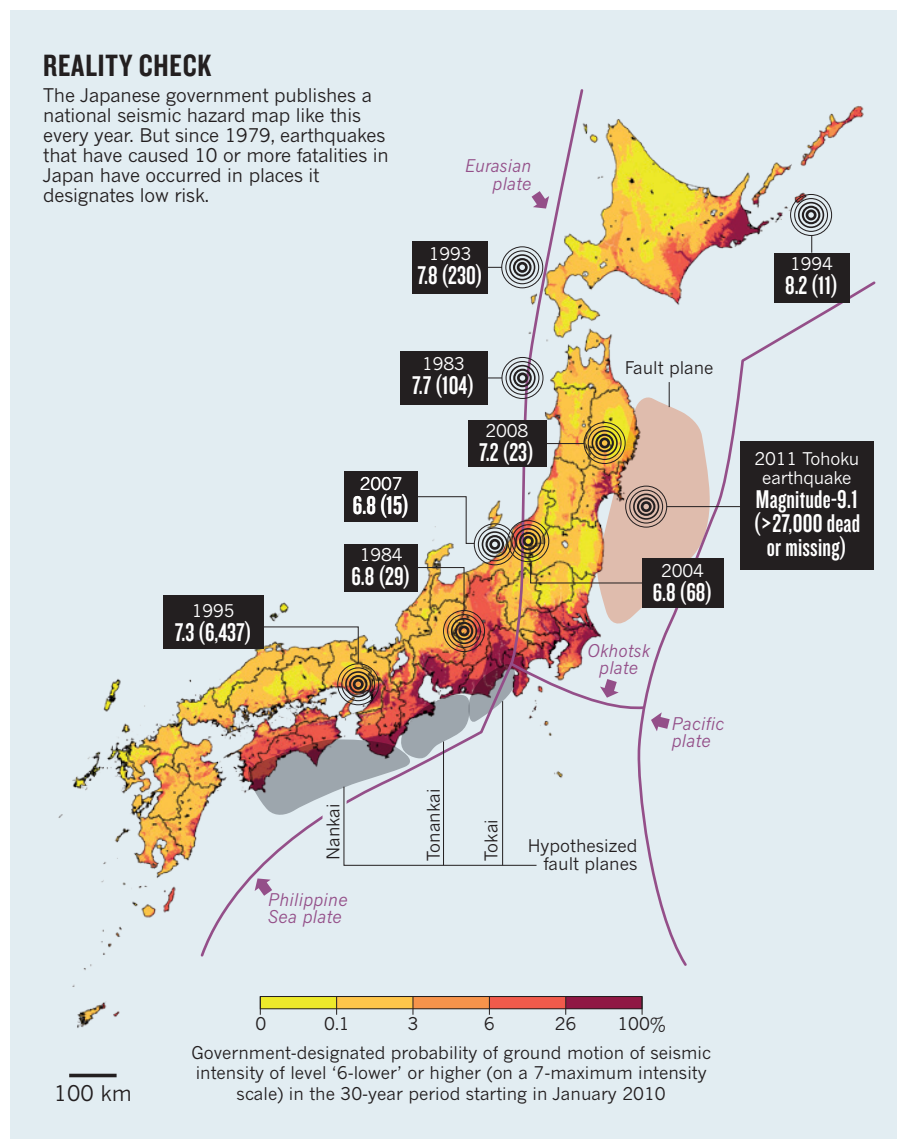
UNPREDICTABLE EARTHQUAKES
Throughout most of seismological history, the prediction of earthquakes hours or days in advance has, for good reason, been regarded with great scepticism[8] (see go.nature.com/ahc6nx). However, in the late 1960s and early 1970s, several studies, initially by researchers in the Soviet Union, and followed by similarly positive studies from major US institutions, led to a burst of optimism. The editors of *Nature* wrote in 1973 that the "situation is in some ways similar to that in 1939 when nuclear fission suddenly became a reality"[9]. Positive results were also published at roughly the same

FUKUSHIMA CRISIS
➲ WWW.NATURE.COM/JAPANQUAKE

REALITY CHECK

The Japanese government publishes a national seismic hazard map like this every year. But since 1979, earthquakes that have caused 10 or more fatalities in Japan have occurred in places it designates low risk.

Eurasian plate

1993 — 7.8 (230)
1994 — 8.2 (11)
1983 — 7.7 (104)
Fault plane
2008 — 7.2 (23)
2007 — 6.8 (15)
2011 Tohoku earthquake — Magnitude-9.1 (>27,000 dead or missing)
1984 — 6.8 (29)
2004 — 6.8 (68)
1995 — 7.3 (6,437)
Okhotsk plate
Pacific plate
Nankai / Tonankai / Tokai
Hypothesized fault planes
Philippine Sea plate

0 0.1 3 6 26 100%

Government-designated probability of ground motion of seismic intensity of level '6-lower' or higher (on a 7-maximum intensity scale) in the 30-year period starting in January 2010

100 km

time in *Science* and some leading speciality journals.

The positive reports were based on claims to have observed 'precursors' of earthquakes. For example, some studies of the type discussed in *Nature*'s 1973 article claimed to have observed decreases of 10–20% in crustal seismic velocities before earthquakes, with the return of the velocities to their normal values being the sign that an earthquake was imminent. But the 1976 earthquake in Tangshan, China, which caused a reported 240,000 fatalities, was not predicted, and by the late 1970s it had become clear to most researchers that the supposed precursors were artefacts. The prediction boom then largely died out, but like many similar examples (such as polywater and cold fusion), die-hard holdouts in several countries continue to make precursor claims.

BASELESS PREDICTION LAW

By the mid-1970s, public discussion of the supposedly imminent Tokai earthquake reached quasi-panic levels. This was exploited by the Japan Meteorological Agency (JMA) and university scientists, who persuaded the Japanese parliament to enact the Large-Scale Earthquake Countermeasures Act (LECA) in 1978. This law in effect requires the JMA to operate a 24/7 monitoring system to detect precursors indicating that the 'Tokai earthquake' (see map) will occur within up to three days. If and when signals thought to be precursors are ever observed, a panel of five geophysicists will review the data, the JMA director will inform the prime minister, and the cabinet will then declare a state of emergency, which will stop almost all activity in a wide area around the Tokai district.

This law, which has no precedent in any other country, presumes of course that reliable precursors exist. In particular, on the basis of one report of a geodetic precursor of an earthquake in Japan in 1944 (see Fig. 2 in ref. 6), geodetic slip is the main target of the JMA observations. The 1944 data, taken far from the epicentral region, were interpreted as possibly suggesting uplift of a few centimetres due to slow slip on a deep part of the fault shortly before the main shock. Unfortunately, the data were measured using antiquated surveying techniques, and are subject to considerable uncertainty. Nothing of this type has ever been observed using Global Positioning System devices or other modern measurement techniques. A famous report of a supposed geodetic precursor, the 'Palmdale
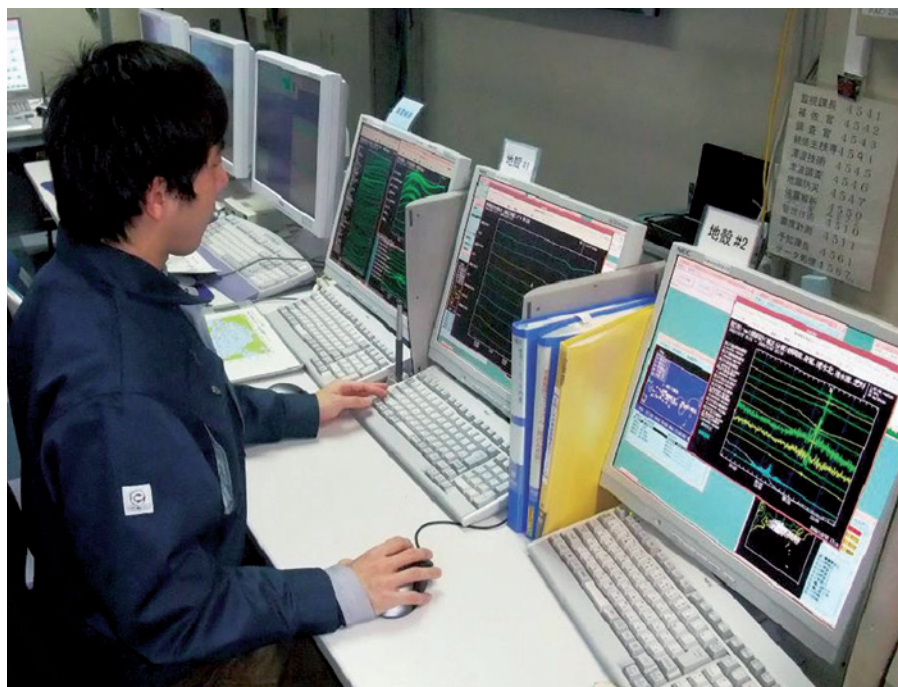
The Japan Meteorological Agency control room conducts monitoring to predict the 'Tokai earthquake'.

Bulge', in the United States in the 1970s was later shown to be an artefact[8].

Basing even a large-scale programme of observational research on the 1944 data would be uncalled for. It beggars belief, then, that the Japanese government operates a legally binding earthquake-prediction system on this basis. The JMA's official home page says (author's translation): "At present the only place a system for predicting earthquakes exists is for a magnitude-8 earthquake with an epicenter offshore Suruga Bay, i.e. the 'Tokai earthquake'. Science and technology have not progressed sufficiently to allow other earthquakes to be predicted." But there are many more observatories now than in 1978. If it really were possible to predict the 'Tokai earthquake' then, surely it would be possible to predict all magnitude-8 earthquakes now.

TIME FOR OPENNESS

How is it that the Tokai prediction system has been in place for more than 30 years, with barely a whimper from most mainstream Japanese seismologists? The reasons for this silence are complex. First, many researchers have been co-opted in various ways (such as with funding and committee memberships). Second, government decisions are nominally reviewed, but review panels are chosen by bureaucrats of the agency being reviewed. Third, cogent criticisms do get reported by print media, but are usually ignored by broadcasters, so critics don't get much traction. Fourth, through the 'press club' system, the government pipes its views directly into the media, often through reporters lacking in

scientific knowledge. Finally, as long as the LECA stays on the books, the government can claim that it is obligated by law to try to predict the Tokai earthquake.

It is time to tell the public frankly that earthquakes cannot be predicted, to scrap the Tokai prediction system and to repeal the LECA. All of Japan is at risk from earthquakes, and the present state of seismological science does not allow us to reliably differentiate the risk level in particular geographic areas. We should instead tell the public and the government to 'prepare for the unexpected'[10] and do our best to communicate both what we know and what we do not. And future basic research in seismology must be soundly based on physics, impartially reviewed, and be led by Japan's top scientists rather than by faceless bureaucrats. ∎

Robert J. Geller *is in the Department of Earth and Planetary Science, Graduate School of Science, University of Tokyo, Tokyo 113-0033, Japan.*
e–mail: bob@eps.s.u-tokyo.ac.jp

1. Headquarters for Earthquake Research Promotion, National Seismic Hazard Maps for Japan (2005); available at http://go.nature.com/yw5e92.
2. Kagan, Y. Y. & Jackson, D. D. *J. Geophys. Res.* **96,** 21419–21431 (1991).
3. Kagan, Y. Y. *Bull. Seism. Soc. Am.* **86,** 274–285 (1996).
4. Newman, A., Schneider, J., Stein, S. & Mendez, A. *Seismol. Res. Lett.* **72,** 647–663 (2001).
5. McCaffrey, R. *Geology* **36,** 163–266 (2008).
6. Ando, M. *Tectonophysics* **25,** 69–85 (1975).
7. Ando, M. *Tectonophysics* **27,** 119–140 (1975).
8. Geller, R. J. *Geophys. J. Int.* **131,** 425–450 (1997).
9. *Nature* **245,** 174 (1973).
10. Kanamori, H. *Seismol. Res. Lett.* **66**(1), 7–8 (1995).

> "We should instead tell the public and the government to 'prepare for the unexpected'."

Call for 'liberation': a November 2010 rally in Edmonton, Canada, for a multiple-sclerosis treatment.

The rise of people power

Calls in Canada for trials of a contentious treatment for multiple sclerosis illustrate how social media can affect research priorities, say **Roger Chafe** and his colleagues.

In 2008, Paolo Zamboni, a vascular surgeon from the University of Ferrara in Italy, hypothesized a new cause of multiple sclerosis (MS). He pointed to abnormalities in the veins draining the brain and spinal cord, a condition he called chronic cerebrospinal venous insufficiency (CCSVI). Zamboni proposed that unblocking the veins by mechanically widening them can improve the symptoms of the disease. He termed his treatment 'the liberation procedure'.

Zamboni's hypothesis is a radical departure from the prevailing view that MS is mainly an autoimmune disorder. In most countries, it has received little attention. In Canada, however, a national debate is under way about whether publicly funded clinical trials should be conducted on the treatment of CCSVI, or even whether patients with MS should have immediate, publicly funded access to a vein-widening procedure called venoplasty. This is despite the fact that virtually none of the country's MS physicians and researchers, and not even the Multiple Sclerosis Society of Canada — the nation's largest support organization for patients — have advocated for either.

What is fascinating here is the extraordinary mobilizing power of the media and the Internet. More than 500 Facebook groups, pages and events dedicated to promoting vascular treatment for MS have emerged in less than two years and amassed tens of thousands of participants. Indeed, the case indicates the unprecedented pressures that politicians and funders worldwide can now face to alter research priorities even in the absence of credible scientific evidence.

In this new social-media environment, researchers and clinicians need to engage more actively with the public to articulate the science validating, or debunking, novel treatments — and to ensure that patients' concerns and priorities are heard.

HIGH HOPES

MS is a chronic neurological disease that causes a variety of symptoms — including weakness, loss of balance or vision, and memory loss — and can be functionally disabling. In the 'relapsing–remitting' form of the disease, symptoms can improve for months or years before worsening again. Currently there is no cure.

Zamboni proposed that obstructions in veins lead to a build-up of iron deposits in the central nervous system, which triggers an autoimmune response[1]. In a non-randomized, non-blinded study of 65 patients, he found that those with the relapsing–remitting form of the disease had fewer relapses after venoplasty. Patients with progressive forms of MS showed little improvement[2]. On the basis of these findings, Zamboni has called for randomized trials to assess the effects of venoplasty more rigorously.

Most neurologists and other physicians who treat patients with MS say that a non-randomized, non-blinded trial is poor evidence that venoplasty is beneficial, especially given the variable nature of some forms of the disease. Moreover, several studies[3,4] have failed to replicate Zamboni's original findings. Nonetheless, in Canada, CCSVI has garnered an extraordinary amount of attention.

In November 2009, *The Globe and Mail* — one of the country's leading national newspapers — and the Canadian Television Network's news programme *W5*, featured stories about Zamboni and CCSVI. Both stories described patients with MS, including Zamboni's wife, as experiencing dramatic improvements after venoplasty. The *W5* programme described it as "a revolutionary treatment for a most debilitating disease [that] could free MS patients from a lifetime of suffering"[5]. Reports in the Canadian media about Zamboni and 'the liberation procedure' have appeared almost weekly since.

The reluctance of the Multiple Sclerosis Society of Canada, clinicians and researchers to advocate for patients' immediate access to venoplasty or even for clinical trials to test its efficacy has led hundreds of patients and their supporters to form advocacy organizations. These groups have used the Internet and social media to share information and testimonials about positive responses to the therapy, to attack the credibility of those advocating caution and to organize campaigns and demonstrations. Some have even accused the MS society and MS physicians of being swayed by conflicts of interest as "when CCSVI is introduced, the number of MS patients will drop".

Partly in response to pressure from such groups, Canada's largest public funder of health research, the Canadian Institutes of Health Research, in partnership with the MS society, convened an expert panel in August last year to assess the CCSVI–MS hypothesis. The panel concluded that more observational studies — for instance, comparing the frequency of abnormal venous blood flow in people with MS with that in individuals who do not have the disease — should be performed, but "in the absence of clear and convincing evidence for CCSVI, the performance of an interventional venous angioplasty trial with its attendant risk to MS

E. KAISER; EDMONTONJOURNAL.COM

patients is not appropriate at this time"[6].

Yet the issue rumbles on. Whether a clinical trial that subjects patients to an invasive and potentially risky procedure ought to be conducted was one of the main focuses of the 2010 annual meeting of the country's federal and provincial health ministers. In December last year, Michael Ignatieff, the leader of Canada's largest federal opposition party, stated his support for clinical trials of endovascular treatment of MS. Meanwhile, hundreds of Canadian patients are travelling, at their own expense, to receive treatment from private clinics around the world now offering venoplasty and in some cases stenting (in which a small tube is placed in a vein to help it stay open) to treat CCSVI.

Although severe adverse events from these procedures are rare, at least two patients have reportedly died after receiving treatment[7]. Complications include clotting, serious bleeding, stents moving to a different part of the body and abnormal heart rhythms.

FACEBOOK EQUIPOISE

What lessons should scientists take from the Canadian CCSVI experience?

A recent poll found that about half of Canadians are now familiar with the use of venoplasty to treat MS. Canada does have a higher prevalence of MS than many other countries. But seemingly unique to Canada are the initial extremely positive news stories about Zamboni and CCSVI, the effective use of social media by patients, and ongoing media interest. We searched 2 leading newspapers in each of 7 countries between November 2009 and January 2011, and found 83 articles on Zamboni and CCSVI in Canadian papers, 16 in Italian papers, and 6 articles in total in the newspapers from the other 5 countries (see 'Hot topic').

Unproven treatments have long been proposed, and tried, for diseases. Tools such as Facebook and YouTube make it considerably more likely that patients will learn about such therapies, without necessarily learning about their potential limitations. Similarly, the mobilizing power of social networking means that unprecedented pressures can be applied to politicians and research funders to expand access to procedures for which there is little scientific support.

A clear lesson is that the traditional approaches for communicating scientific findings to the public and to policy advisers (reports, briefing notes, press releases and news conferences) are insufficient. When patient groups are using social media to advocate and mobilize, scientists must use similarly effective tools to communicate.

Scientists and research funders also need to avoid adopting an 'us versus them' mentality. Many researchers have been frustrated by the lack of impact that science often has on public debates, but new models for engagement are emerging. For instance, the James Lind Alliance in Oxford, UK, brings together patients and clinicians to identify questions that have not received adequate attention from industry or academia. The patients and clinicians then jointly identify priorities for research.

A thornier issue is whether patient pressure and social media should affect the design and timing of clinical studies.

The dominant paradigm is that an interventional trial is not justified unless there is a strong biological rationale supported by observational studies, and 'clinical equipoise' regarding its efficacy — that is, a lack of consensus between researchers and clinicians about whether the treatment is preferable to others that are already available. This strategy is logical and defensible. But in today's era of 'Facebook equipoise', it may make sense in rare cases to conduct a clinical trial before the desired weight of scientific evidence

accumulates; for instance, if thousands of patients are exposing themselves to risks and costs of unevaluated medical procedures.

The Multiple Sclerosis Society of Canada and the National Multiple Sclerosis Society in the United States are funding seven studies to evaluate the association between CCSVI and MS. If these studies consistently show no association, this may be enough to convince most people with MS that venoplasty is not worth trying. If the results are delayed, or are unconvincing to many patients, we believe that the benefits of a double-blinded randomized trial would outweigh the costs — including the diversion of resources from other priorities. Such a study would involve giving one group of patients a vein-widening procedure and another a sham procedure, and having the assessor being unaware of which patients received which procedure.

> *"Traditional approaches for communicating scientific findings to the public are insufficient."*

Rigorous evaluation has previously been important in helping to refute unproven medical claims, such as the purported benefits of the drug Laetrile (amygdalin) for cancer[8], or the alleged dangers of the measles, mumps, rubella (MMR) vaccine in causing autism.
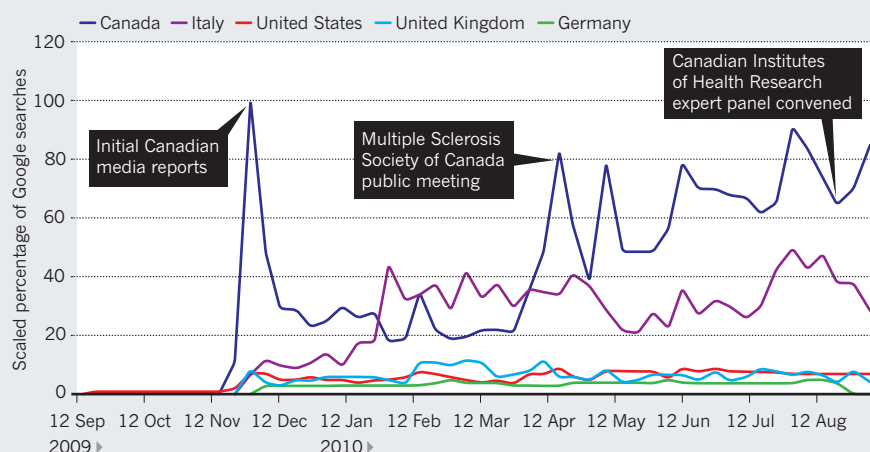
In the long term, to prevent an increasing proportion of public resources being diverted to testing what will probably turn out to be ineffective or harmful therapies, more effort needs to be devoted to improving the scientific literacy of the public, politicians and the media — and to engaging with a public that is no longer deferential to experts. ∎

Roger Chafe *is director of paediatric research at the Memorial University of Newfoundland, St John's, Newfoundland and Labrador A1B 3V6, Canada.* **Karen B. Born** *is in the Department of Health Policy, Management and Evaluation at the University of Toronto, Toronto, Ontario M5T 3M6, Canada.* **Arthur S. Slutsky** *and* **Andreas Laupacis** *are at St Michael's Hospital, Toronto, Ontario M5B 1W8, Canada, and King Saud University, Riyadh, Saudi Arabia.*
e-mail: roger.chafe@med.mun.ca.

1. Zamboni, P. *et al. J. Neurol. Neurosurg. Psychiatry* **80,** 392–399 (2009).
2. Zamboni, P. *et al. J. Vasc. Surg.* **50,** 1348–1358 (2009).
3. Sundström, P. *et al. Ann. Neurol.* **68,** 255–259 (2010).
4. Doepp, F., Paul, F., Valdueza, J. M., Schmierer, K. & Schreiber, S. J. *Ann. Neurol.* **68,** 173–183 (2010).
5. CTV News. *The Liberation Treatment: A whole new approach to MS* (21 November 2009); available at http://go.nature.com/krl2o7.
6. Canadian Institute of Health Research *Joint Invitational Meeting on Multiple Sclerosis Research — Summary Report* (2010); available at http://go.nature.com/drf2l5.
7. Rudick, R. A. *Nature Rev. Neurol.* **6,** 472–474 (2010).
8. Lerner, I. J. *Cancer* **53,** 815–819 (1984).

HOT TOPIC

A measure of the number of Google searches for the term CCSVI (chronic cerebrospinal venous insufficiency) reveals that public interest in Canada soared in just one year.

Canada — Italy — United States — United Kingdom — Germany

Initial Canadian media reports

Multiple Sclerosis Society of Canada public meeting

Canadian Institutes of Health Research expert panel convened

y-axis: Scaled percentage of Google searches — 0, 20, 40, 60, 80, 100, 120

x-axis: 12 Sep 2009, 12 Oct, 12 Nov, 12 Dec, 12 Jan 2010, 12 Feb, 12 Mar, 12 Apr, 12 May, 12 Jun, 12 Jul, 12 Aug

Greenland's ice cap, pictured here in a photo by US clothing retailer and research patron Gary Comer, serves as a vast climate-change archive.

Lands' End to the Arctic

Henry Pollack relishes a climate-science narrative with an intrepid and passionate businessman at its heart.

Two narratives make up the fabric of *The Fate of Greenland*. The polar-research perspectives of well-known climate scientists form the main thread, woven through with a posthumous tribute to Gary Comer. Comer, founder of US clothing manufacturer Lands' End, became a patron of climate science in his later years, in a most unusual way.

A native of Chicago, Illinois, Comer had as a young man crewed on sailing vessels large and small, developing a passion for remote places. In its early days, Lands' End — founded in 1963 — reflected his abiding love of sailing, offering sails, lines, hardware and nautical apparel for day-sailors on Lake Michigan and owners of luxurious ocean-going yachts. Three decades later, Comer's business success enabled him to resume exploration of the world with his own ship, *Turmoil*, a modified and well-appointed 46-metre North Sea trawler.

Comer took *Turmoil* to some extraordinary places. In 1997, he envisioned a trip to the Russian Far East

to retrace the path of Vitus Bering in his 1728 voyage through the Bering Strait into the Arctic Ocean. Comer invited naturalist Philip Conkling and me to join him on that voyage, which began, as did Bering's, in Kamchatka, the massive volcanic peninsula that drops south like a giant pendulum from easternmost Russia. We saw much of the same terrain that Bering did, and visited the island, now called Bering Island, where he perished after his ship was wrecked there on his second voyage.

Although we failed to pass through the Bering Strait in 1997, the Arctic Ocean did not elude Comer for long. In 2001 he again headed north with *Turmoil*, this time along

The Fate of Greenland: Lessons from Abrupt Climate Change
PHILIP W. CONKLING, RICHARD ALLEY, WALLACE BROECKER & GEORGE DENTON, WITH PHOTOGRAPHS BY GARY COMER
MIT Press: 2011. 224 pp. £22.95/$29.95

the east coast of North America, to enter the Northwest Passage at its eastern portal in Baffin Bay. Sixteen days later, *Turmoil* emerged into the Beaufort Sea, north of Alaska, the first private vessel in modern times to traverse the Northwest Passage unassisted by an icebreaker. Roughly a century earlier, it had taken Roald Amundsen three years to make that voyage.

In an interview in the *Chicago Tribune Magazine* shortly after this historic 2001 traverse, Comer said: "All along the way we were astonished by how easy and quick our voyage was compared to who had gone before us. We began to realize that the Arctic...is the first part of the world where wildlife and humankind are experiencing changes brought on by global warming, and these changes will be difficult, if not impossible, to cope with."

Through this, Comer became very interested in climate science. Disappointed with the pace of research funded through normal channels, he wanted to accelerate climate studies. He sought out creative scientists and provided both funds and logistical support via his ship and its on-board reconnaissance float-plane. Over the next four years he took

G. COMER

climatologists Richard Alley, Wally Broecker, and George Denton to Greenland and the Arctic, along with Conkling, the chronicler of earlier Comer expeditions.

Cancer overtook Comer in 2006, ending a remarkable life that had embraced the beauty of the natural world and fostered a great passion for protecting and preserving it. *The Fate of Greenland* records his voyages of scientific exploration in the Arctic, through the words of Alley, Broecker, Denton and Conkling and Comer's own remarkable photographs.

The importance of Greenland in the global climate system, now and in the past, cannot be overstated. It influences polar meteorology, hosts seasonal sea ice on its periphery, and through ice discharge affects the thermohaline circulation of the adjacent Atlantic Ocean and the hydrological cycle around the world. The Greenland ice sheet, 3 kilometres thick in places and second only to the accumulation on Antarctica, has been an archive of climate change in the Arctic for more than 100,000 years. Much of what we know about events such as the Younger Dryas, the mid-Holocene Optimum, the Medieval Warm Period and the Little Ice Age derives from Greenland. And today, it is undergoing rapid changes.

For those actively engaged in climate science, there will be little new in this book. But its charm and value lies in the informal narratives these scientists provide into the ways that scientific enquiry proceeds: the posing of important questions, the formulation of many hypotheses, the quick rejection of most, and the structuring and execution of tests to evaluate the remaining few. Alley's work on the history of Arctic climate reconstructed from the deep ice cores, Broecker's discussion of the decades-long evolution of his thinking about thermohaline circulation, and Denton's descriptions of his on-the-ground experience in glaciated landscapes are stitched together by Conkling in an almost seamless narrative that preserves their individual voices.

The final chapters ask what lessons Greenland has taught us. The documentation of abrupt climate change gleaned from its ice offers a sober warning of the consequences of a rapidly warming Arctic, including the summertime loss of Arctic sea ice, the accelerating thawing of the permafrost surrounding the Arctic Ocean, and other issues such as methane release and global sea-level rise. The authors recognize that some mitigation of climate change remains possible, if the political will exists, but note that adaptation to the effects of climate change has already become a necessity — one that looms larger with every policy delay. ∎

Henry Pollack *is a professor of geophysics at the University of Michigan, Ann Arbor 48109-1005, USA, and author of* A World Without Ice *(Avery/Penguin, 2009). e-mail: hpollack@umich.edu*

G. H. H. HUEY/CORBIS

Eradicating some species to protect others, such as California's island fox, can stir controversy.

Reconstituted Edens

T. C. Boyle's latest novel probes the convoluted impacts of species eradication programmes, finds **Emma Marris**.

Islands are like novels. They are self-contained worlds, populated by a manageable cast of characters. And most have narratives that hinge on a series of incidents, accidents, coincidences, births and deaths.

Species arrive by flying, rafting, swimming — and by luck or by accident. The order in which they show up determines much. Early colonists can claim ecological niches and put down roots. But in limited spaces, existence is tenuous. One bad year, and every member of a species may die. Thus, on small islands, as in novels, the death of a single individual can change the trajectory of the whole.

T. C. Boyle's new novel, *When the Killing's Done,* centres on an archipelago off the coast of California called the Channel Islands. The plot mirrors real campaigns there to remove introduced animals in a bid to protect native species.

In 2001, rats were poisoned on Anacapa Island by the National Park Service to prevent the rodents from eating native birds' eggs. Pigs were eradicated from Santa Cruz Island a few years later to protect indigenous flora and deter golden eagles — lured to the island to gorge on pork — from enjoying a side dish of endangered island fox.

Not everyone was behind the push to restore the archipelago to its pre-human glory. In 2003, *The Washington Post* told of animal-rights activist Rob Puddicombe scattering "vitamin-fortified kibble" around Anacapa in a bid to help the rats there survive — vitamin K being a partial antidote to the rodenticide that was used. His act was in

When the Killing's Done
T. C. BOYLE
Viking: 2011. 369 pp. $26.95
Bloomsbury: 2011, 384pp. £18.99

vain. Today Anacapa is rat-free.

Puddicombe explained his actions to the reporter. "To me, the idea of species is just an abstract concept. Species go extinct all the time," he said. "These animals are here and alive now. Their lives have value."

Boyle uses the conundrum of killing individuals to save species as the central conflict of his novel. He counterpoises a fictionalized version of Puddicombe, David LaJoy, with a fictional National Park Service employee in charge of eradications, Alma Boyd Takesue. Add mother–daughter relationships, shipwreck, deaths and Boyle's trademark detailed descriptions of characters deciding to have a drink, and you have the book.

The philosophical questions are not explored in the depth that some might hope for. The protagonists only briefly reflect on their opposing positions, and the reasons they came to their beliefs are not fleshed out. LaJoy and Takesue are presented as equally intransigent, equally misanthropic, equally angry and, for me, equally unlikeable.

These two remain more or less ideologically static throughout the book, but plenty happens. There's science, crime, pig-hunting, sheep-farming and accidents — lots of accidents. Boyle's characters run afoul of the forces of nature nearly as often as they decide to crack open one more cold beer. "People fall from cliffs, people drown, people get drunk and do violence to one another, bones break, hearts give out, and it's all in a day's work for the Park Service," muses Takasue as she watches tourists clomp all over Santa Cruz.

Some accidents bring people to the islands; others end in death, just as nature's accidents and extinctions determine what species we consider "native" to particular islands. Near the novel's end, Boyle brings a new species to Santa Cruz, raising intriguing hypothetical questions. If a new creature appeared on the island, would scientists assume that humans had imported it and summarily remove it? Or would they leave it alone? And does it matter how it got there? Is there any sense in which ecologists can not meddle with the islands they take care of?

The characters in *When the Killing's Done* are in mourning for a simpler past. "How much better would it be if nobody ever came out here and the islands could exist in the way they always had. Or should have," thinks Takasue. But, as always, we showed up. Island biogeography may be a matter of accident, but humanity is the inevitable mishap. ∎

Emma Marris *is a writer based in Columbia, Missouri, USA. Her book on the future of conservation,* Rambunctious Garden, *comes out in September. e-mail: e.marris@gmail.com*

Books in brief
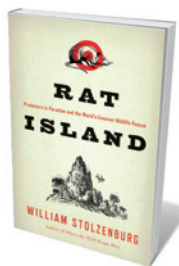
Rat Island: Predators in Paradise and the World's Greatest Wildlife Rescue
William Stolzenburg BLOOMSBURY *288 pp. $26 (2011)*
Sometimes you have to be cruel to be kind. So say conservationists who eradicate invasive species in order to protect more established ones. Focusing on one island in the Bering Sea where nesting birds are being savaged by marauding rats, journalist William Stolzenburg explains why ecologists have become poisoners to reduce the rodent population. He highlights the fine line between wildlife protection and poaching.
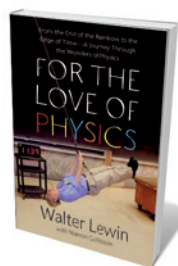
For the Love of Physics: From the End of the Rainbow to the Edge Of Time — A Journey Through the Wonders of Physics
Walter Lewin with Warren Goldstein SIMON & SCHUSTER *320 pp. $26 (2011)*
Having topped the online video charts with his daredevil YouTube lectures, physicist Walter Lewin has become a celebrity teacher. In this book, he brings his trademark hands-on approach to explaining the wonder of physics. By asking simple questions that we can all identify with, such as why we are shorter standing than lying down, he explains complicated subjects such as gravity.
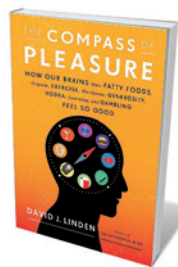
The Compass of Pleasure: How Our Brains Make Fatty Foods, Orgasm, Exercise, Marijuana, Generosity, Vodka, Learning, and Gambling Feel So Good
David J. Linden VIKING *240 pp. $26.95 (2011)*
The pursuit of desire explains much of human history. Neuroscientist David Linden explains how aspects of our brain lead us to crave food, sex, drugs and rock'n'roll. He also notes the evolutionary reasons why we gain pleasure from virtues as well as vices. Knowledge of the biological basis of pleasure forces us to rethink the moral and legal ramifications of addictions, he says.
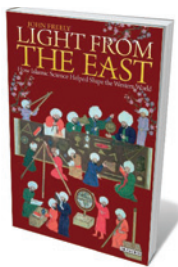
Light From the East: How the Science of Medieval Islam Helped to Shape the Western World
John Freely I. B. TAURIS *256 pp. $28 (2011)*
In his account of the rich history of idea-swapping among medieval civilizations, travel writer John Freely celebrates the golden age of Arabic science. By describing the astrologers, physicians, philosophers, mathematicians and alchemists of the Muslim world, he reveals how knowledge spread from Central Asia and the Middle East to Spain and Europe, where it influenced Western thinkers.
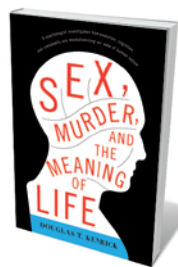
Sex, Murder, and the Meaning of Life: A Psychologist Investigates How Evolution, Cognition, and Complexity are Revolutionizing Our View of Human Nature
Douglas Kenrick BASIC BOOKS *256 pp. $26.99 (2011)*
With heads full of sexual — and sometimes homicidal — fantasies, we humans still carry much evolutionary baggage from our wild ancestors. In his idiosyncratic account, social psychologist Douglas Kenrick shows how these roots are more beneficial than troubling. Apparently, irrational and selfish behaviours — such as one-night stands, prejudices and greed — are ultimately rational, he says, and presage positive urges such as love and generosity.

S. TORFINN/PANOS

Psychologists are trying to understand the brutal behaviour this Rwandan genocide survivor endured.

The empathy gap

Malfunctioning brain networks only partly explain why some people act cruelly, finds **Stephanie Preston**.

In his 2007 book *Musicophilia*, psychiatrist Oliver Sacks warned that although neuroscience offers exciting insights, "there is always a certain danger that the simple art of observation may be lost, that clinical description may become perfunctory, and the richness of the human context ignored". Simon Baron-Cohen, director of the Autism Research Centre in Cambridge, UK, rises to the challenge in his latest book by combining basic science and clinical observation in an attempt to explain human cruelty.

In *Zero Degrees of Empathy*, Baron-Cohen reconstrues "evil" as the product of a failure to empathize, caused by malfunction in an empathy network within the brain. Most of the book focuses on detailed case studies of disorders of empathy: psychopathic personality disorder (Type P), borderline personality disorder (Type B), narcissistic personality disorder (Type N), and the autism spectrum disorders (ASD). Filling the gap that Sachs notes, Baron-Cohen's rich descriptions of the patients' social and emotional lives force us to rethink common beliefs about the nature of these disorders.

He splits the four types into two forms of "zero degrees of empathy". Types P, B, and N are dubbed "Zero Negative" because their lack of empathy is considered to be solely detrimental. Individuals with ASD are "Zero Positive", as they lack empathy but also have "systematizing" natures, which are valued because their focus on detail and patterns are key components of advanced processes such as hypothesis testing.

Baron-Cohen then attributes these traits to dysfunction in the brain's "empathy circuit", genetic abnormalities and poor developmental conditions. These preclude the forging of an emotionally-resilient "internal pot of gold" — a robust psychology that is normally developed through secure attachment.

Baron-Cohen's introduction of these abstract terms is understandable, since they allow a complex topic to be understood by a broad audience. They reflect some truth about the world, yet allow him to incorporate his empirical work into a broad social theme. And he should be lauded for approaching the problem head-on, which is rarely done in neuroscience.

However, his terminology also makes it difficult to judge exactly how his model advances our understanding of human cruelty. Empathy was already understood as a (statistically) normally distributed phenomenon that arises in interconnected neural regions that are dysfunctional in these disorders. Terms such as "empathy circuit" may lead people to incorrectly infer that these neural regions are solely associated with empathy, when in reality they participate in a range of domains including language, action selection, decision-making, emotion, attention, and general social behaviour.

Nevertheless, the disturbing examples of cruel behaviour in the book, including the seemingly gratuitous levels of humiliation of victims in genocides and massacres, should be used to inform our scientific theories. Humiliation can be used to establish status, to signal collaboration with a dominant person or to respond to one's own perceived oppression. Additionally, dehumanization may arise when perpetrators have to side-step their intact empathy mechanisms in order to permit murder via indirect methods — such as henchmen, technology or by forcing victims to kill one another.

An interdisciplinary framework that combines our neuroscientific knowledge with findings from social and political science may allow us to capture the 'richness of the human context' in such a consequential topic.

Understanding our simultaneous capacity for great compassion and cruelty is no easy feat. We should take Baron-Cohen's accessible book as an invitation to leave the comforts of our smaller, more tractable problems in a genuine attempt to address larger social issues. ∎

Zero Degrees of Empathy: A New Theory of Human Cruelty/The Science of Evil: On Empathy and the Origins of Cruelty
SIMON BARON-COHEN
Allen Lane/Basic Books: 2011.
208 pp./256 pp.
£20/$25.99

Stephanie Preston *is assistant professor in cognition and cognitive neuroscience, University of Michigan, Ann Arbor, 48109-1043 USA.*
e-mail: prestos@umich.edu

ART

Blurring the boundaries

A touring exhibition shows the value of the Swiss 'artists-in-labs' programme, find **Deborah Dixon**, **Harriet Hawkins** and **Mrill Ingram**.

With sci–art collaborations on the rise, this refreshing exhibition stands out as a reflection on how both parties can benefit from the tensions and challenges in such meetings of minds. *Think Art — Act Science* is the first exhibition to showcase art that has come out of the Swiss 'artists-in-labs' programme, a joint project of the Institute for Cultural Studies in the Arts at the Zurich University of the Arts and the federal culture ministry.

Since 2003, 28 artists, most from Switzerland, have been placed in labs there and, since 2009, in China. This exhibition of 'prototypes' by nine of them looks at how interdisciplinary collaborations work, and how the ideas they spark can best be conveyed to the public. Showing at Arts Santa Monica in Barcelona, Spain, until 15 May, the exhibition then tours several US cities and Ireland.

Some installations capture the artists' visceral response to their scientific engagements. Sylvia Hostettler's polyester resin sculptures (*Regeneration I* and *II*, 2008), their protuberances glowing with malign intensity, suggest the monstrous nature of the phenomena she encountered in the Center for Integrative Genomics at the University of Lausanne. Equally unsettling are Pe Lang's *Kinetic Speakers*, a cluster of sleek, swivelling microphone-cum-loudspeakers that simultaneously listen and react to sound. These finely crafted electronic sculptures, made in collaboration with the Swiss Center for Electronics and Microtechnology, Alpnach, dispel notions of media technologies as passive.

That the works are prototypes is significant — the programme stresses the collaborative production of ideas over a finished output. Each work is accompanied by a video in which the artists and scientists convey the ways in which communication did and did not take place, and discuss how the works evolved. These videos reveal how all involved had their expectations shaken.

Scientists at the Swiss Federal Institute of Aquatic Science and Technology in Dübendorf, for example, which hosted artist Ping Qui, expressed surprise that her work, like their research, privileges precision and rigour. In her *Breathing Plants* (2008), tumescent plastic sculptures resembling roots grow and shrink in response to the humidity in a room, accompanied by recorded sounds of human inhalation and exhalation.

Also on display is Christian Gonzenbach's light sculpture *QUARC, Quantum Art Crystal* (2009), which emerged from his residency at the University of Geneva physics department and CERN, where he came to admire the aesthetic sensibility and craft of scientists there.

A response to experiments on dark matter and complex systems, *QUARC* deploys a crystalline lattice of neon tubes that flash on and off in response to unpredictable movements of a series of small magnets. Meanwhile, the video of scientists at the Artificial Intelligence Laboratory, University of Zurich, shows them enthused by the challenge of creating the dancing robot envisioned by choreographer Pablo Ventura.

The videos touch on often-overlooked aspects of collaboration, such as alienation and miscommunication. Wenfeng Liao, for instance, former artist-in-residence at the Swiss Federal Institute of Forest, Snow and Landscape Research in Birmensdorf, spoke of finding lab science strange. His point of engagement was the researchers' office plants. His photo *Office Plants* (2010), for instance, which shows snapshots of the plants nailed to the trees the scientists study, led to a discussion about artificial boundaries between living and working environments and those serving as research subjects.

This contemplative exhibition demonstrates, with considerable style, the value of sci–art collaboration in opening up productive spaces of mutual engagement, inspiration and intrigue. ∎

Think Art — Act Science
Arts Santa Monica, Barcelona, Spain, until 15 May 2011.

↻ **NATURE.COM**
Martin Kemp on neuroscience and art:
go.nature.com/dskqe2

Deborah Dixon *and* **Harriet Hawkins** *are based at the Institute of Geography and Earth Sciences, Aberystwyth University, UK.* **Mrill Ingram** *is based at the Department of Geography, University of Wisconsin-Madison, USA.*
e-mail: hah7@aber.ac.uk

Small Paths (2), a video installation by Wenfeng Liao, former artist-in-residence at the Swiss Federal Institute of Forest, Snow and Landscape Research.

CORRESPONDENCE

'No' to ban on stem-cell patents

The advocate-general of the European Court of Justice has recommended the prohibition on ethical grounds of patents involving human embryonic stem cells (Nature 471, 280; 2011). We write to express profound concern over this recommendation, as coordinators of multinational European stem-cell projects, working with both adult and embryonic stem cells.

Embryonic stem cells are cell lines, not embryos. They are derived using surplus in vitro fertilized eggs donated after fertility treatment and can be maintained indefinitely. As more than 100 established lines are now supplied through national and international cell banks, concern about commercialization of the human embryo is misplaced.

It is premature to suggest that human embryonic stem cells can be replaced in development of therapies. Although induced pluripotent stem cells offer additional possibilities, particularly for disease modelling, the reprogramming process is still imperfect.

Scientists working in stem-cell medicine will not be able to deliver clinical benefits without the involvement of biological industry. But innovative companies must have patent protection as an incentive to become active in Europe.

The advocate-general's opinion therefore represents a blow to years of effort to derive biomedical applications from embryonic stem cells in areas such as drug development and cell-replacement therapy. If implemented, European discoveries could be translated into applications elsewhere, at a potential cost to the European citizen. The advocate-general's opinion will now be considered by members of the court. We trust that they will deliberate on the full implications before making a legally binding ruling.

Austin Smith on behalf of 12 co-signatories*, Wellcome Trust Centre for Stem Cell Research, Cambridge, UK.
Austin.Smith@cscr.cam.ac.uk
*A full list of signatories is available online at go.nature. com/ryt4jv. Three authors declare competing financial interests (see go.nature.com/ryt4jv).

'Plan B' for X-ray astronomy

The European Space Agency (ESA) decision to go ahead without NASA support in selecting its next large astronomy mission (Nature 471, 421; 2011) has dashed hopes for the International X-ray Observatory (IXO). X-ray astronomy needs a cheaper 'plan B'.

NASA might yet contribute 20% of the funding for a smaller IXO, but that would not be comparable as a flagship for NASA, ESA and the Japan Aerospace Exploration Agency (JAXA).

A radical rethink of IXO might do it. A mission is defined by its science, not its hardware. IXO spanned all three decades of the X-ray band with a single mirror. But X-ray optics need focal lengths proportional to the photon energy reflected, so each sub-band would benefit from having a very different mirror. Turning IXO into a programme of independent small satellites, each optimized for one band, could be the answer.

ESA, NASA and JAXA could each choose one and build it, without financial or technological interdependence, and space the launches over several years.

The X-ray sky is full of complex sources with much interesting detail, but satellites dedicated to different energy bands would aim mostly at point sources, which have no discernable structure. Hence a further satellite with a high angular resolution imager would be needed.

Some inter-agency agreement would help to achieve synergy, coordinating where they should be looking and allowing for multi-telescope proposals.

Martin Elvis Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts, USA.
martinselvis2@gmail.com

Bio-PIN could cut biobanking risks

Donors of specimens to biobanks are entitled to respect for their privacy and protection from discrimination (Nature 469, 156–157; 470, 169–70; 471, 159–160; 2011). So long as the identity of a donor is on record, such protection cannot be guaranteed. A new system uses a biological PIN code to get around this problem. By keeping biological specimens and data anonymous, the 'Bio-PIN' also encourages donors by offering superior protection (J. J. Nietfeld et al. Nature Rev. Cancer 11, 303–308; 2011).

Unique to each donor, the Bio-PIN is based on one of their biological characteristics, such as a DNA fingerprint. It means that genetic or other data lodged in a biobank cannot be misused and that privacy is fully protected. As in financial banking, it allows two-way communication between bank and donor.

J. J. Nietfeld University Medical Center Utrecht, the Netherlands.
J.J.Nietfeld@umcutrecht.nl
Jan-Eric Litton Karolinska Institute, Stockholm, Sweden.
J. J. N. declares competing financial interests (see go.nature. com/adinec).

Arab world needs its science diaspora

As members of the Society for the Advancement of Science and Technology in the Arab World, we strongly believe that engaging the scientific diaspora — the research community of Arab scientists working abroad — is crucial for successful science in the region.

In addition to the underlying political and societal problems and failed economic policies responsible for recent dramatic events in Arab countries, the failure of the educational systems is to blame. Academic institutions are not producing the skilled workforce necessary to meet local challenges and compete in global economies (Nature 470, 147–149; 2011). The domino effect of the protests and change we are witnessing today, and its potential consequences for regional stability, indicate that correcting this shortcoming should be a regional and international priority.

The international community must, through focused research and funding initiatives, assist in the areas of education, research and technology likely to have the highest impact on society. Arab scientists abroad represent a huge reservoir of talent for catalysing these efforts and strengthening partnerships with countries in the region. Investing in training and research programmes that harness the expertise, resources, networks and enthusiasm of these Arab scientists would contribute significantly to reversing the 'brain drain'. And it would encourage local efforts to establish academic and research institutions to nurture creativity and entrepreneurship.

Hilal A. Lashuel Swiss Federal Institute of Technology, Lausanne (EPFL), Switzerland.
hilal.lashuel@epfl.ch
Wael K. Al-Delaimy University of California, San Diego, USA.

EARTH SCIENCE

Lithosphere today …

Seismic images of the Colorado plateau region reveal a mantle 'drip' forming under the Grand Canyon area. This hidden process may be responsible for the puzzling uplift of the plateau. SEE LETTER P.461

GEORGE ZANDT & PETER REINERS

Large-scale uplift, erosion and volcanic activity are standard features of mountain belts that are subject to the horizontal tectonic forces associated with plate movements. But somehow, in the past 70 million years, possibly quite recently, the virtually undeformed Colorado plateau of the southwestern United States was uplifted about 2 kilometres, intruded by diverse magmas, and eroded and deeply incised, creating a dramatic topography that includes the 1.8-km-deep Grand Canyon. This enigmatic history has perplexed geologists for more than a century.

On page 461 of this issue, Levander et al.[1] provide a possible explanation that relates uplift, erosion and volcanism on the surface to deep features on the underside of the plateau, where it sits atop the convecting upper mantle. Understanding these features and their surface manifestations is a timely challenge as we begin to recognize the critical part that deeper mantle processes must play in the evolution of the continents.

Using new seismic data from the Earthscope Transportable Array, the authors combine different seismic techniques to get a better view of the lithospheric structure of the Colorado plateau. Continental lithosphere is the strong, long-lived crust and upper mantle, extending to about 150 km depth and comprising less-dense crustal rocks down to about 40 km and denser mantle rocks below that. This lithospheric package sits on the convecting asthenosphere, the part of the mantle below the lithosphere that is hotter, weaker and perhaps less dense — a potentially unstable configuration.

The seismic images reveal an anomalous region in the lower crust and upper mantle under the Grand Canyon and underlying much of the western half of the plateau (Fig. 1). In this region, the mantle part of the lithosphere seems to form a vertically elongated viscous drip sinking into the Earth, pulling off the lower part of the crust above it, a process called convective instability[2] or delamination[3]. Once the drip detaches and is replaced by upwelling hot asthenosphere, the remaining crust above it bobs up and is often intruded by magmas. This deep process may be manifesting
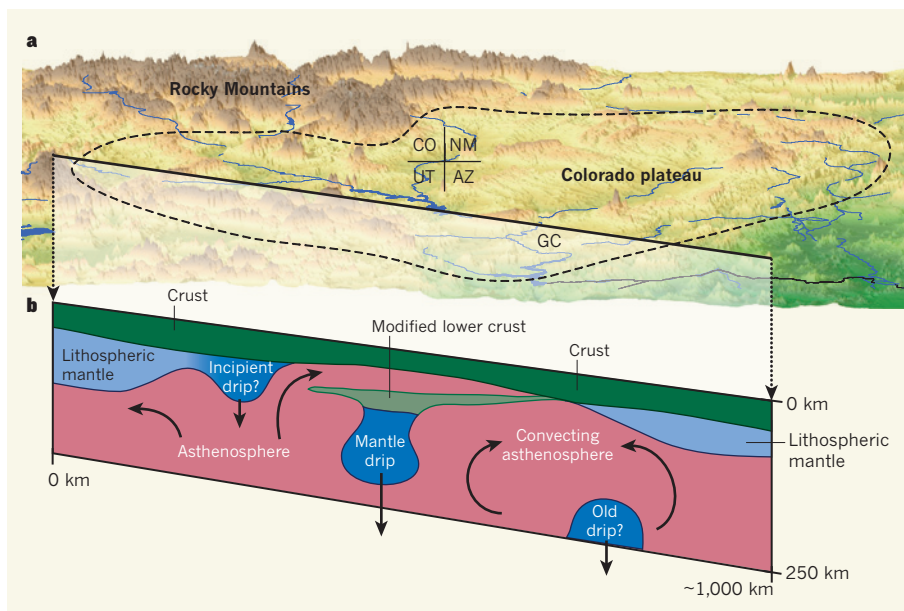
Figure 1 | Over and under the Colorado plateau and surrounding regions. **a**, Low-angle topographic view of the area, which lies within the states of Colorado, New Mexico, Arizona and Utah. Dashed line, the approximate location of the plateau's physiographic boundary; blue, rivers, lakes and reservoirs; GC, Grand Canyon. **b**, Cross-section of the lithospheric structure along the transect shown in **a**. Levander et al.[1] used multiple seismic imaging techniques to identify an active mantle drip peeling off the lower crust over a large area centred near the Grand Canyon. This subsurface process may be responsible for geologically recent uplift in the area. A series of such drips, occurring over tens of millions of years, may account for the regional uplift of the entire Colorado plateau.

itself at the surface through uplift, volcanism and erosion in and around the Grand Canyon. Levander et al.[1] hypothesize that this lithospheric drip formed in the past 6 million years and is just the latest of several such events that occurred around the edges of the Colorado plateau, producing rock and surface uplift over the past 20 million to 30 million years.

Drips are turning up in many places. In the western United States alone, geophysicists also see them under the Sierra Nevada[4], Wallowa Mountains[5] and central Great Basin[6]. They seem to be telling us something important about the upper mantle and how it affects the surface of continents. But exactly what they are telling us, other than that the upper mantle is often drippy, requires the integrated perspective offered by deep seismic imaging and clues to the palaeoelevation and erosion of the surface through time.

The work[1] is supported by recent studies of

volcanic rocks and mantle xenoliths — rocks carried by magmas from as deep as 120 km below the surface. Volcanism seems to have been encroaching on the plateau for the past 25 million years or so, and the geochemistry of the volcanic rocks shows increasing contributions from the convecting asthenosphere, rather than the ancient lithospheric mantle beneath the crust[7]. Levander and colleagues' main observation of one major lithospheric drip is thus superimposed on the longer-term picture of progressive erosion of the lithosphere by asthenospheric convection on the edges[8]. Decreased strength and increased density leading to dripping of the formerly buoyant lithosphere are thought to have resulted from its viscous weakening by hydration[9] from fluids released by an unusually flat subducting slab, followed by infiltration by relatively dense, iron-rich melts from the asthenosphere after removal of the flat slab about 25 million years

The creation of lawrencium, the eleventh new man-made element beyond uranium (element 92), was announced on April 13 by the University of California and the U.S. Atomic Energy Commission. An isotope of this element was formed on February 14 by bombarding … californium (element 98) with boron-10 or boron-11 nuclei in a heavy-ion linear accelerator at the Lawrence Radiation Laboratory of the University of California; the four nuclear scientists who carried out this experiment … propose to name the new element lawrencium in honour of the late Ernest Orlando Lawrence, inventor of the cyclotron and founder of the laboratory which bears his name … Although it may be possible to create one or two more new elements by this technique, it is improbable that elements beyond number 105 will be synthesized and satisfactorily identified.
From *Nature* 29 April 1961

100 Years Ago

Last summer a pair of robins built their nest in an old fish-basket that was hanging in a shed at the back of my house. All went well until the young birds were about a week old — then happened what appeared to me to be a catastrophe. My Aberdeen terrier pup "Bebe", who must have had some natural desire to catch the mother bird, managed one morning to make a meal of her. Contrary to what I should have expected, the male bird kept close to his young family. Day by day I turned over part of the garden to supply him with a little help in his task. In due course he taught the whole of his young family to fly. I have made enquiries, but cannot find anyone who has had a similar experience, and wondered what your readers might know about such cases.
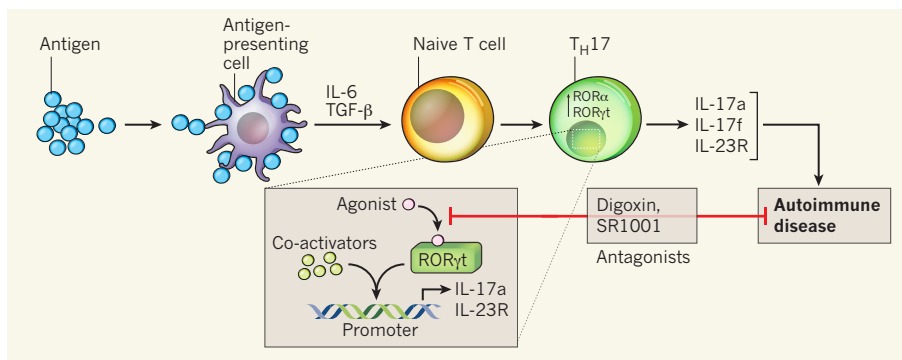**From *Nature* 27 April 1911**

Figure 1 | Effects of digoxin and SR1001 on T_H17 cells. On encountering an antigen on the surface of antigen-presenting cells (and in the presence of IL-6 and TGF-β), naive T cells differentiate into T_H17 cells. This event is associated with expression of the nuclear receptors RORγt and RORα. These receptors, particularly RORγt, are required for T_H17-cell differentiation and for the expression of IL-23R and IL-17a, among other cytokines. Two studies[1,2] show that digoxin and SR1001 bind RORγt, possibly by competing with the natural agonists of these receptors. By inhibiting the recruitment of co-activators and promoting the recruitment of co-repressors, these antagonists reduce RORγt transcriptional activity, T_H17-cell differentiation and IL-17 production, and delay the onset and reduce the severity of autoimmune disease in mice.

of genes encoding IL-17 and the IL-23 receptor (IL23R). The digoxin-induced changes in gene-expression profiles are similar to those observed in RORγ-deficient cells[1], a finding consistent with the notion that digoxin exerts its effects by inhibiting RORγt activity. Previous studies[3,5] demonstrated that expression of either RORα or RORγ in T cells induces the expression of IL-17a. Huh *et al.* show that digoxin inhibits RORγ-dependent, but not RORα-dependent, induction of IL-17a. This result is consistent with the RORγ specificity of digoxin.

Treatment with digoxin or SR1001 greatly inhibited the expression of messenger RNAs for IL23R, IL-17a, IL-17f and IL-22, and markedly reduced production of the IL-17a protein[1,2]. RORγt regulates IL-17a and IL23R expression directly by binding to promoter sequences of the genes encoding these proteins[1,5] (Fig. 1). Treatment with digoxin or SR1001 significantly reduced the binding of RORγt to these sequences. These observations are consistent with the idea that the antagonists' binding causes a conformational change in the ligand-binding domain of the receptors that negatively influences their interaction with co-activators and promotes co-repressor recruitment.

SR1001 showed no obvious toxicity at the doses tested. Digoxin, however, was toxic for human cells at concentrations lower than those needed to inhibit RORγt. Huh *et al.*[1] therefore synthesized digoxin derivatives that retained the RORγt-antagonistic effects but were much less toxic in human cells. Intriguingly, in addition to inhibiting T_H17-cell differentiation, these derivatives increased the expression of IFN-γ and FOXP3 in human CD4+ T cells; these are markers of two other T-cell types, T_H1 and T_{reg} cells, respectively. This finding suggests that inhibiting RORγt activity also promotes the differentiation of human naive T cells into other effector-cell lineages. By contrast, neither

digoxin nor SR1001 affected the differentiation of mouse naive T cells into other lineages[1,2].

In mice, loss of RORγ greatly reduces the development of experimental autoimmune encephalomyelitis[3,5]. Both teams[1,2] demonstrated that treatment with either digoxin or SR1001 delays the onset of this disorder in mice and reduces its severity. This was associated with a reduction in the number of T_H17 cells entering the animals' spinal cord. The investigators therefore propose that RORγt antagonists might be effective for treating autoimmune diseases. But first a number of caveats must be considered.

Apart from its expression in T_H17 cells, RORγ is expressed in several other cell types and tissues, in which its function is unknown. It is therefore unclear what side effects long-term treatment might induce in these tissues. Moreover, as recently outlined[7], the role of T_H17 cells and their associated cytokines is complex. So, although inhibiting RORγt may have therapeutic merit for autoimmune disease, it might adversely affect the beneficial functions of these cells in fighting pathogens. Despite these concerns, however, generating more-potent and more-selective derivatives of digoxin and SR1001 could offer attractive strategies for treating autoimmune disorders. ∎

Anton M. Jetten *is in the Cell Biology Section, Division of Intramural Research, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina 27709, USA.*
e-mail: jetten@niehs.nih.gov

1. Huh, J. R. *et al. Nature* **472,** 486–490 (2011).
2. Solt, L. A. *et al. Nature* **472,** 491–494 (2011).
3. Ivanov, I. I. *et al. Cell* **126,** 1121–1133 (2006).
4. Jetten, A. M. *Nucl. Recept. Signal.* **7,** e003 (2009).
5. Yang, X. O. *et al. Immunity* **28,** 29–39 (2008).
6. Kumar, N. *et al. Mol. Pharmacol.* **77,** 228–236 (2010).
7. O'Connor, W. Jr, Zenewicz, L. A. & Flavell, R. A. *Nature Immunol.* **11,** 471–476 (2010).

MATERIALS SCIENCE

Colour without colourants

Creating coloured polymer films without the use of pigments might seem impossible. But using miniature polymer spheres, and a novel assembly process, this feat has been accomplished over large film areas.

PAUL V. BRAUN

Humans have been intensely interested in colour for tens of thousands of years. Until the advent of modern organic chemistry, the colours found in art, decorations and clothing were obtained from pigments, for example indigo, cochineal and ochre, which are derived from naturally occurring compounds found respectively in plants, insects and earth. Amazingly, until about 50 years ago we did not realize that the natural world also makes extensive use of structural colour — colour formed not by the interaction of light with a dye, but by the diffraction of light from a material containing a periodic or quasiperiodic structure on the length scale of the wavelength of visible light. Writing in *Advanced Materials*, Finlayson *et al.*[1] describe a method for creating intensely coloured polymer films based on the principles of structural colour used by nature.

The iridescence of mother-of-pearl, the shimmer of an opal, the striking coloration of many beetles and the brilliant colours of peacock feathers (Fig. 1a) — these are all due to structural colour, not pigment molecules[2]. For example, the coloration of an opal comes exclusively from its microstructure, which consists of a periodic arrangement of silica spheres embedded in a silica-based matrix that has a slightly different refractive index from that of the spheres. As different wavelengths of light interact with this structure, they are diffracted in different directions, leading to the iridescence with which we are all familiar.

Over the past two decades, scientists have become exceptionally good at creating materials exhibiting strong colorations that are based on the principles of structural colour found in nature. These 'artificial opals', or more generally — when not made of colloidal building blocks — photonic crystals, have been proposed as potentially forming the basis of a wide range of technologies, ranging from chemical sensors to on-chip optical waveguides[3]. It has been extremely difficult, however, to form structurally coloured materials over the large areas required for applications such as decorative coatings and fabrics. Finlayson *et al.*[1] now show that highly scalable processing techniques can be used to create coloured, large-area polymer films based on the principles of structural colour (Fig. 1b).

The coloured films reported by the authors derive their hue from the same principle as the coloration of an opal, and are in fact also based on spheres. In their case, the specially designed polymer-based spheres, each with a diameter of about 250 nanometres, form the films when merged together. Until recently, artificial opals tended to have rather muted colours, with flashes of iridescence at specific angles but overall a rather milky and unappealing appearance. But a few years ago, researchers discovered[4] that the colour could be made much more intense by the addition of a fraction of a per cent of carbon black. At that time, however, it was not clear how these materials could be made into the large-area structures required for many applications.

Until Finlayson and colleagues' study, the artificial opals with the largest areas were created through the technique of spin coating[5], in which a solution is placed on a rotating substrate to form a thin film. Spin coating works well for coating flat surfaces of up to about a metre across with an opalescent film, but becomes entirely impractical either for coating a curved surface or for making free-standing films. Opalescent films can also be formed by convective assembly processes, although these struggle to coat surfaces much larger than a glass slide[6], and by other generally slow self-assembly processes based on particle interactions in solution. The authors now report[1] their use of edge shearing, a potentially rapid and industrially scalable process in which colloidal particles are sandwiched as a thin film between two removable polymer sheets and drawn over a sharp edge. The edge-shearing process induces the particles to form a highly periodic structure, greatly strengthening their response to light and making their colour much more intense.

In addition to this novel colloidal assembly process, Finlayson *et al.* used colloidal particles that are different from normal. They consist of a hard, crosslinked polystyrene core surrounded by a soft polyethylacrylate shell. Between the core and shell is a thin polymer layer that binds the core and shell together. When heated and pressed, the particles' soft shells merge, forming a continuous structure, with the cores remaining intact and separated. Once the cores have been induced to assemble into a periodic structure, through the edge-shearing process, a strong iridescence results that is due to the difference in the refractive indices of the core and the shell.

The edge-shearing process as demonstrated by the authors is rather slow (about 1 centimetre per second), and there may be challenges to scaling up the synthesis of the core–shell colloidal particles. But through appropriate engineering these shortcomings are almost certainly resolvable. A more interesting question is whether the films can be made multispectral — that is, with more than one colour at once. The current films exhibit highly saturated colours that vary as a function of the viewing angle. Thus, they can provide a brilliant rainbow of colours, but they cannot present two colours concurrently from the same viewing angle.

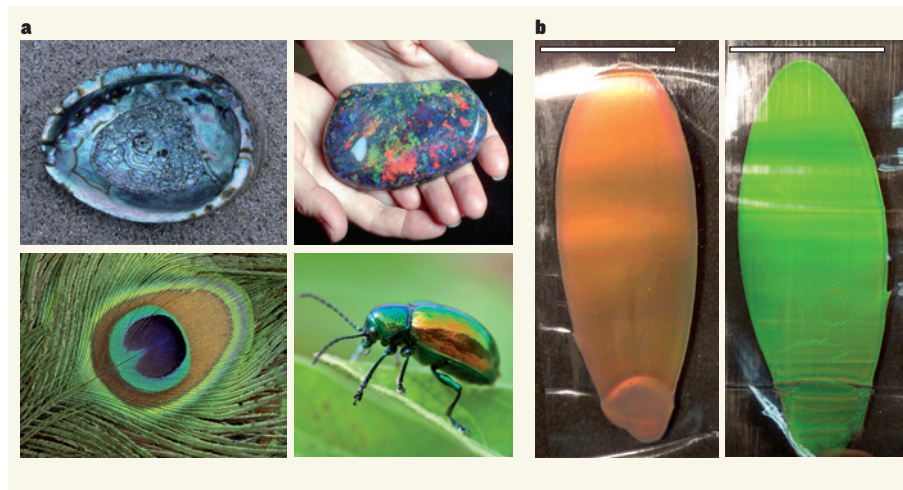Might it be possible to self-assemble structures that contain two characteristic length

Figure 1 | Pigment-free colour. a, Structural colour in nature (clockwise from top left): mother-of-pearl, an opal, a beetle and a peacock feather. **b**, Finlayson *et al.*[1] created these brilliant polymer films by edge-induced rotational shearing of a solid assembly of core–shell polymer particles. Scale bars, 3 centimetres.

scales, and thus colours, at the same time? Of course, other colours could be added using dye molecules, but such an approach would not harness the power of the technology. Still, even with this limitation, an exciting range of possibilities emerges, both for consumers and for industry. Imagine a car whose colour changes as it drives by, or fabrics that shimmer in an unexpected way. Looking further, it is likely that the structures could be engineered to be responsive and change colour as a function of their environment, providing both aesthetically and technologically important capabilities. ■

Paul V. Braun is in the Department of Materials Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA.
e-mail: pbraun@illinois.edu

1. Finlayson, C. E. et al. Adv. Mater. **23,** 1540–1544 (2011).
2. Kinoshita, S. Structural Colors in the Realm of Nature (World Scientific, 2008).
3. Arpin, K. A. et al. Adv. Mater. **22,** 1084–1101 (2010).
4. Pursiainen, O. L. J. et al. Appl. Phys. Lett. **87,** 101902 (2005).
5. Jiang, P. & McFarland, M. J. J. Am. Chem. Soc. **126,** 13778–13786 (2004).
6. Jiang, P. et al. Chem. Mater. **11,** 2132–2140 (1999).

ANIMAL BEHAVIOUR

Large–scale cooperation

A deeper understanding of the evolution of cooperation will come from investigations of what animals know about working together. A study with Asian elephants now adds to the literature on the subject.

AMANDA M. SEED & KEITH JENSEN

We are cooperating right now. You are sharing in a communicative venture that is the product of two authors coordinating their views on a scientific joint effort, mediated by an editor and a publishing team. This is no small feat; it is the result of the evolution of complex cognition that might be uniquely human. But there is increasing evidence that the evolutionary roots of cognitively rich cooperation may be found in our closest relatives, chimpanzees and bonobos. In research reported in the *Proceedings of the National Academy of Sciences*[1], Plotnik *et al.* propose that elephants (which are only very distantly related to humans) can cooperate as cleverly as these great apes. Such an independently evolved case could help us identify the evolutionary pressures that shape intelligent minds.

Plotnik and colleagues required pairs of elephants to work together using a procedure developed by primate researchers[2], scaled to elephantine proportions. A single rope was threaded through a platform baited with food. Pulling one end with its trunk resulted in the elephant getting the rope but not the platform. It took two elephants pulling together to succeed.

Plotnik *et al.* did not stop there — nor did the elephants. The animals had to learn to wait for their partner when they were not released together. And if the partner had no access to the rope, the test subject did not pull at all. Two of the elephants found their own solutions to the problem posed by Plotnik's team, one by placing her foot on her rope end and letting her partner pull, the other by waiting at the starting gate for his partner before approaching the platform.

Plotnik *et al.* conclude that elephants (Fig. 1), like chimpanzees[2,3] and unlike birds[4], may understand the role of the partner in a cooperative enterprise. Elephants (and most chimpanzees) did not perform perfectly: at first, when their partner was delayed, they pulled their end of the rope regardless. Similarly, rooks (a member of the crow family) spontaneously pulled in the platform in pairs, but did not wait for their partner. But when the delay was built in gradually (a procedure not used with the birds), both chimpanzees and elephants quickly learned not to pull until the partner was there, and in the case of the elephants, only when the partner could reach his end of the rope.

This overt cooperation need not, however, require any special social cognition. Instead,

Figure 1 | Togetherness. Asian elephants may understand the role of a partner in a cooperative enterprise.

the elephants might have learned to pick up on the cues in the environment that meant pulling would be worthwhile (the sight of another elephant with the rope 'in-trunk', or the feeling of tension in the rope), just as we might wait for a traffic light and coordinate our behaviour with it to our advantage. The elephants certainly made strikingly few errors, but they live in a conservation centre and have been trained to perform log-hauling, painting and even music-making for visitors. As such, they are trained to be trainable. Future tests, such as seeing how readily the elephants can learn to wait for a non-social cue, such as a green light, are needed to distinguish between what the authors refer to as "a well-developed propensity toward partner-oriented, deliberate cooperation" and a more general propensity to learn quickly.

Nevertheless, elephants and chimpanzees learned to wait with comparable ease, so are the authors right to claim that their cooperative skills are on a par? An important finding yet to be explored in elephants is that chimpanzees bring about the opportunity to cooperate by recruiting their partner, going beyond a learned recognition of favourable conditions.

In the 1930s, Crawford showed[5] that if one chimpanzee was given such a glut of bananas as to be uninterested in the task, the other would attempt to engage her, for example by pulling on her hand. In one more recent study[2], a chimpanzee recruited a human experimenter (but not another chimpanzee) by leading him to the apparatus. In another[3], the researchers taught the chimpanzees to wait, but then required the animals to remove a bolt to release a partner when one was needed. This they did. The chimpanzees had not simply learned to wait to pull until their group mate had the rope in reach, but they had grasped the causal role of the partner. The ability to rapidly transfer knowledge learned in one context to perform a new behaviour is a useful test of cognitive flexibility.

Chimpanzees seem to know enough about their group mates to be able to use them as social tools. There is no evidence yet that elephants do. However, this might reflect the alien nature, to a human observer, of elephant communication. Plotnik and colleagues

stress that it will take time to decode their low-pitched calls and subtle positional cues. Another new study[6] describes how elephants led by mature females, but not younger ones, bunch around their young when they hear a male lion's roar. But these authors could not discern how the elephants coordinated their defensive stance.

Plotnik et al.[1] have broadened our appreciation of how different animals cooperate. Testing the world's largest land animal was a mammoth undertaking, but by adapting experimental techniques used on other species, notably primates, they have allowed direct comparisons to be made. To see how evolution shapes minds for cooperation, future work should compare closely related species that have different social systems. Perhaps a propensity to 'tune-in' to another's behaviour is an adaptation to social living more generally, and need not require large brains. Even some fish, with their small brains, are able to coordinate their hunting efforts[7]. Further research on various creatures great and small will help us learn about how animals — including humans — understand and coordinate with each other. ■

Amanda M. Seed is in the School of Psychology, University of St Andrews, St Andrews, Fife KY16 9JP, UK. Keith Jensen is in the Research Centre for Psychology, School of Biological and Chemical Sciences, Queen Mary, University of London, Mile End Road, London E1 4NS, UK.
e-mails: ams18@st-andrews.ac.uk; k.jensen@qmul.ac.uk

1. Plotnik, J. M., Lair, R., Suphachoksahakun, W. & de Waal, F. B. M. Proc. Natl Acad. Sci. USA 108, 5116–5121 (2011).
2. Hirata, S. & Fuwa, K. Primates 48, 13–21 (2007).
3. Melis, A. P., Hare, B. & Tomasello, M. Science 311, 1297–1300 (2006).
4. Seed, A. M., Clayton, N. S. & Emery, N. J. Proc. R. Soc. B 275, 1421–1429 (2008).
5. Crawford, M. P. Comp. Psychol. Monogr. 14, 1–88 (1937).
6. McComb, K. et al. Proc. R. Soc. B doi:10.1098/rspb.2011.0168 (2011).
7. Bshary, R., Hohner, A., Ait-el-Djoudi, K. & Fricke, H. PLoS Biol. 4, e431 (2006).

MATERIALS CHEMISTRY

Polymer networks take a bow

A new study reports that the shapes and surface patterns of thin films of a stretched material can be modified by shining ultraviolet light at it. The resulting topologies depend on the exposure pattern, the applied stress and the sample thickness.

WILHELM T. S. HUCK

For applications such as tissue engineering, drug-delivery systems and biosensors, there is a need for stimulus-responsive polymers that can adapt to their environment by converting optical, thermal or mechanical signals into chemical signals, and vice versa[1]. Writing in *Advanced Materials*, Bowman and colleagues[2] report just such a material, and demonstrate a process in which thin films of the material respond to light and mechanical forces by forming intricate surface patterns.

In traditional polymer networks, crosslinks are used to connect polymer chains. Well-known examples are rubbers (elastomers) and many of the gels that are often used in biomedical applications. Elastomers show large deformations in response to stretching, but only the flexible polymer segments between two crosslinks are affected during deformation — they adopt less coiled, more linear conformations. The crosslinks themselves remain fixed.

Bowman and colleagues[2], however, exploit the unique properties of an intriguing class of polymeric material known as a covalent adaptable network[3] (Fig. 1a). In these materials, the covalent bonds that hold polymer networks together form reversibly, and fall apart if appropriate external stimuli are supplied. This means that the bonds can be repeatedly and controllably rearranged. Several chemical reactions can be applied to make suitably reversible covalent bonds in these materials[3], but Bowman and co-workers used the thiol–ene reaction, in which a thiol compound (R–SH, where R is a hydrocarbon group) is added to a carbon–carbon double bond through a mechanism involving free radicals.

Having prepared their elastomeric polymer network, the authors added a photoinitiator — a compound that generates radicals when exposed to ultraviolet light. These radicals were a crucial part of the adaptable network: once formed, they triggered a reversible 'addition–fragmentation chain transfer' (AFCT) reaction[4]. The net result of this reaction was to shuffle the connectivity of the network around, breaking and remaking the chemical crosslinks holding the elastomer together.
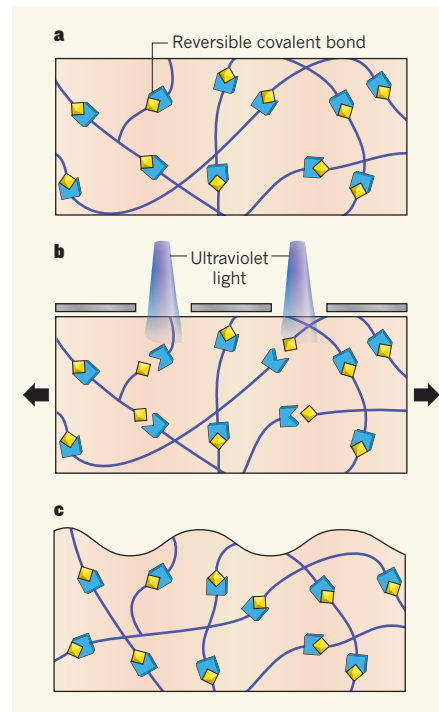

Figure 1 | Mechanophotopatterning. a, Covalent adaptable networks consist of crosslinked polymer chains. The crosslinks contain reversible covalent bonds that break apart under the influence of a specific external stimulus. **b**, Bowman and colleagues[2] stretched samples of a light-responsive covalent adaptable network and shone ultraviolet light at them through masks, destroying crosslinks in the irradiated areas. **c**, In response to the mechanical stress, material in the irradiated areas flowed towards the non-exposed regions, altering the surface topology of the samples. New crosslinks then formed, fixing the altered topology into place.

Bowman and colleagues went on to study what happens if AFCT occurs at selected regions while the elastomer is stretched. Under these conditions, the covalent crosslinks in the affected areas break apart, leaving two reactive ends (Fig. 1b). The authors reasoned that this would cause a sharp drop in the density of crosslinks in those areas, so that the material would turn into a viscous fluid that undergoes plastic flow when the mechanical stress is applied — in much the same way that wet chewing gum flows when stretched. The flow would release the stress until the material reached a new equilibrated morphology, whereupon the reactive ends could find each other again and form new crosslinks, freezing the altered structure in place (Fig. 1c).

The authors found that a thin film of their polymer did indeed behave in this way. When they shone ultraviolet light through a mask onto a stretched thin film of the polymer, the resulting plastic flow caused the material to move away from the light-exposed regions. Overall, the surface of the film developed little bumps of material in areas that remained dark, and shallow depressions in the irradiated regions. This demonstrated that a combination

of light irradiation and mechanical deformation can be used to precisely control the topology of light-responsive elastomers, establishing a new technique that Bowman and colleagues dubbed[2] mechanophotopatterning (MPP). Remarkably, the intrinsic material properties of the polymer network remained unchanged after deformation, because the total number of crosslinks was unchanged.

Bowman and colleagues demonstrated that a variety of surface topologies can be produced using MPP, making it potentially useful for many applications. An additional feature of the technique makes it even more broadly useful. When the authors irradiated an optically thick sample of their polymer (a sample through which ultraviolet light could not pass completely), a differential stress was generated between the top and the bottom faces of the sheet. To relieve this stress, the sheet bowed into a three-dimensional shape. Such shapes could have promising applications in the manufacture of lenses for advanced optical systems.

The surface patterns made by the authors[2] using MPP typically contained features that had lateral dimensions of several hundred micrometres and heights of several micrometres, but there is no reason that micrometre-sized features with heights in the tens of nanometres would not form equally well using suitable masks. Whether the definition of MPP-generated features will be comparable to that of features made using alternative patterning techniques, such as those that rely on buckling and creasing to release mechanical stress[5,6], remains to be seen.

Despite the advanced mechanical properties of the material described by Bowman and co-workers, the chemistry involved is straightforward and could be used to make covalent adaptable networks in which the polymer chains have a range of chemical groups attached. I think that the most exciting applications of MPP lie at the interface of cell biology and tissue engineering. One particularly promising application would be to incorporate covalent adaptable networks into flexible scaffolds modified to incorporate ligands that encourage cell binding. Cells are known to exert mechanical forces on substrates to which they bind[7]. Using substrates amenable to MPP would provide an opportunity to dynamically change the mechanical feedback that the substrate gives to a bound cell. This in turn could provide insights into the impact of mechanical forces on cellular function[8], and help in the design of new materials suitable for tissue engineering, especially in cases involving cells that require different mechanical stimuli as a tissue develops[9]. ∎

Wilhelm T. S. Huck *is at the Institute for Molecules and Materials, Radboud University Nijmegen, 6525 AJ Nijmegen, the Netherlands, and in the Department of Chemistry, University of Cambridge, UK.*
e-mail: w.huck@science.ru.nl

1. Stuart, M. A. C. *et al. Nature Mater.* **9,** 101–113 (2010).
2. Kloxin, C. J., Scott, T. F., Park, H. Y. & Bowman, C. N. *Adv. Mater.* doi:10.1002/adma.201100323 (2011).
3. Kloxin, C. J., Scott, T. F., Adzima, B. J. & Bowman, C. N. *Macromolecules* **43,** 2643–2653 (2010).
4. Scott, T. F., Schneider, A. D., Cook, W. D. & Bowman, C. N. *Science* **308,** 1615–1617 (2005).
5. Bowden, N., Huck, W. T. S., Paul, K. E. & Whitesides, G. M. *Appl. Phys. Lett.* **75,** 2557–2559 (1999).
6. Kim, J., Yoon, J. & Hayward, R. C. *Nature Mater.* **9,** 159–164 (2010).
7. Harris, A. K., Wild, P. & Stopak, D. *Science* **208,** 177–179 (1980).
8. Engler, A. J., Sen, S., Sweeney, H. L. & Discher, D. E. *Cell* **126,** 677–689 (2006).
9. Kloxin, A. M., Kasko, A. M., Salinas, C. N. & Anseth, K. S. *Science* **324,** 59–63 (2009).

ASTRONOMY

A new spin on the first stars

Understanding the nature of the first stars, whose formation marked a pivotal epoch in the Universe's history, is at the frontier of astronomy. An analysis of stellar data indicates that they were fast-rotating objects. SEE LETTER P.454

JASON TUMLINSON

Like many of the dinosaurs, the first stars in the Universe were massive beings that died out long ago. On page 454 of this issue, Chiappini *et al.*[1] report measurements of the abundances of heavy elements in stars from one of the Galaxy's oldest star clusters, denoted NGC 6522, that suggest that the earliest generations of stars were not only massive but also rotating rapidly.

Astronomers have made major strides in the study of the first stellar generations using a technique that draws its inspiration from old bones. Massive stars (those at eight times the Sun's mass or more) live fast and die young (after only 30 million years or less), so the first stars that formed within a billion years of the Big Bang are no longer around to illuminate even our largest telescopes. We can read the astronomical 'fossil record' because, when stars die as explosions called supernovae or become red giants, they cast the heavy elements (those heavier than helium) that they have synthesized in their nuclear-burning cores into interstellar space. The heavy elements, which astronomers call metals, then get mixed into fresh gas and are eventually recycled into new stars. Any star formed with roughly the Sun's mass (2×10^{30} kg) or less lives for at least the whole 13.7-billion-year history of the Universe, so small stars bearing the metals produced by the first stars can be studied and the properties of the earlier generations deduced from the relative quantities of each observed element.

By reanalysing data that their group had obtained with the European Southern Observatory's Very Large Telescope, Chiappini *et al.*[1] have measured greatly elevated abundances of the heavy elements strontium (Sr) and yttrium (Y) in eight old stars in NGC 6522, a tightly bound cluster of stars associated with the Galaxy's bulge that is at least 12 billion years old[2] (Fig. 1). The observed enhancements in these two rare elements when compared with iron suggest that the 'slow neutron capture' process for creating elements heavier than iron is up to 10,000 times more effective in the stars that preceded and enriched NGC 6522 than in non-rotating stars of the same mass. Like a spinning vessel filled with water, rotating stars experience mixing between inner and outer nuclear-burning gas layers that would not otherwise overlap. A cascade of nuclear reactions in these overlapping layers generates radioactive neon, which in turn emits a high flux of neutrons that are captured by iron and other heavy nuclei to create rare elements such as Sr and Y. After the death of the original, rapidly rotating stars, these elements eventually found their way into new star-forming clouds and then into the stars in NGC 6522.

To explain the large enhancements seen in Sr and Y, the authors invoke a rotating-star model with a surface velocity of 500 kilometres per second — a dizzying spin compared with the Sun's stately turns at $2 \, \mathrm{km \, s^{-1}}$, or the typical value of $100 \, \mathrm{km \, s^{-1}}$ seen in massive stars in the Milky Way. Indeed, $300–500 \, \mathrm{km \, s^{-1}}$ is rare for Milky Way stars, but could be typical of the first stars. This rapid spin is likely to have affected all stages of stellar formation and evolution in the first generations, from their initial mass to their chemical composition and products. Rapidly rotating first stars may in fact be more likely to die as energetic γ-ray bursts than their slowly spinning cousins, which may make them more easily detectable in the early Universe.

Chiappini and colleagues' results add some observational urgency to a question that has taxed theorists of the first stars in the past few years. Over the past decade, there developed
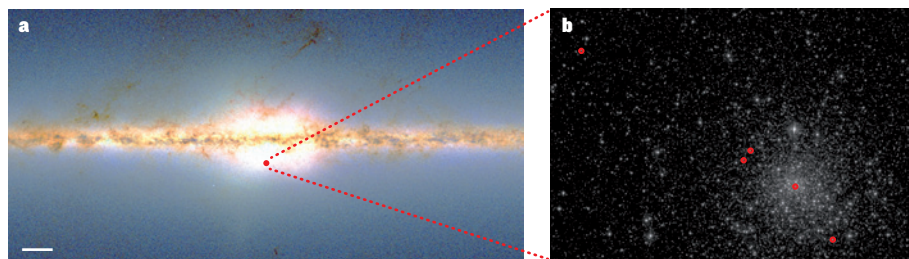
Figure 1 | Zooming in on NGC 6522. Chiappini and colleagues' scrutiny[1] of eight stars in NGC 6522, one of the Milky Way's oldest star clusters, suggests that the Universe's first stars spun rapidly. **a**, Infrared image of the Galaxy's central region; scale bar, 5 degrees on the sky. **b**, Image of part of NGC 6522 obtained by the Hubble Space Telescope/Advanced Camera for Surveys; red circles indicate five of the stars examined by the authors that are covered by the Hubble image.

a tentative theoretical consensus that the first stars formed with typical masses of tens to hundreds of solar masses in the first few hundred million years of cosmic history. The chemical fossil record generally supported this picture. But the exact distribution of masses could not be calculated precisely, because it depended on poorly understood details of the radiation transport and gas chemistry that occur as a forming star's core makes the transition from a dense cloud to stellar densities and begins its nuclear burning. Another wild card in the formation models of the first stars is the total amount of rotation (or angular momentum) in the original cloud from which the star forms: fast rotation could influence the final mass of the star[3] or even promote the cloud's fragmentation, leading to multiple star systems[4,5].

Until now, these models have had no observational constraint on how fast the final star should be rotating, and they could not predict its rotation a priori because they stopped well short of the densities at which stellar evolution and nuclear burning begin. Chiappini *et al.* have provided important guidance to this effort: whatever happens during the complicated final stages of star formation at these early epochs, it must be able to lead to rapidly spinning stars. Indeed, one recent theoretical study[6] suggested that primordial stars can form with both 125 solar masses and at least 800 km s^{-1} of rotation.

Despite these encouraging developments, many questions remain about these eight stars, the cluster in which they reside, and their true relationship to the first stars. Although NGC 6522 has an age that places it within the first billion years of cosmic history, the uncertainty in this age and the cluster's relatively high metal content (one-tenth of the Sun's value) relative to stars at that epoch not associated with clusters (one-hundredth to one-thousandth of the Sun's value) leaves open the possibility that these eight fossils bear the rotation-driven products of stars from several generations after the first.

The rotating-star models used by Chiappini and colleagues to interpret the abundances in NGC 6522 need further development and

independent confirmation by other researchers. Clusters such as NGC 6522 are generally not included in numerical simulations of the first stars and galaxies such as those described above; integrating them is a challenge for theory. Many details remain to be worked out concerning how these rapidly rotating stars form, evolve

throughout their lives and die. Finally, additional abundance signatures of rotating metal-poor stars beyond Sr and Y should be developed and searched for in these stars and others in the fossil record to test the rotating models against multiple lines of evidence. Notwithstanding these open issues, these old stellar fossils will no doubt impart a fresh spin to our thinking about the earliest stars in the Universe. ∎

Jason Tumlinson *is at the Space Telescope Science Institute, Baltimore, Maryland 21218, USA.*
e-mail: tumlinson@stsci.edu

1. Chiappini, C. *et al. Nature* **472,** 454–457 (2011).
2. Barbuy, B. *et al. Astron. Astrophys.* **507,** 405–415 (2009).
3. Tan, J. & McKee, C. F. *Astrophys. J.* **603,** 383–400 (2004).
4. Turk, M. J., Abel, T. & O'Shea, B. W. *Science* **325,** 601–605 (2009).
5. Clark, P. C. *et al. Science* **331,** 1040–1042 (2011).
6. Stacy, A., Bromm, V. & Loeb, A. *Mon. Not. R. Astron. Soc.* (in the press); preprint at http://arxiv.org/abs/1010.0997 (2011).

NEUROSCIENCE

Sleepy neurons?

A study in rats suggests that individual neurons take a nap when the brain is forced to stay awake, and that the basic unit of sleep is the electrical activity of single cortical neurons. SEE ARTICLE P.443

CHRISTOPHER S. COLWELL

Do you yearn for the simple pleasure of getting a good night's sleep and waking up feeling well-rested? Study after study suggests that, for people in industrial or post-industrial societies, one cost of moving away from the farm is the inability to get as much sleep as one desires. This sleeplessness, coupled with the fact that disruptions to sleep and circadian rhythms have wide-ranging, negative consequences for our health, has led to a new level of concern about our sleepless nights. Indeed, many people regularly take sleeping pills — especially the growing population of over-65-year-olds, for some of whom this has become part of the nightly routine[1,2].

To provide relief from these problems, we need to understand more about the mechanisms and functions of sleep. Some mechanistic progress has been made using genetic analyses of mice[3,4] and *Drosophila*[5] — yes, flies do sleep, or at least enter into a rest state that shares many of the core features of mammalian sleep. Still, for the more clinically minded sleep researchers, sleep is defined by distinctive patterns of brain activity, and by the nightly changes in physiology that underlie these brain waves. On page 443 of this issue, Vyazovskiy *et al.*[6] report exciting data suggesting that

sleep patterns can occur in localized regions of the rat brain, and may occur in single cortical neurons even when a rat is awake.

Patterns of neural activity in the brain change across the sleep–wake cycle. These temporal changes can be measured as electric field potentials, and are picked up as 'slow-wave activity' (SWA) using electroencephalography (EEG). Physiologically, neurons in the cortex seem to exist in two stable configurations[7,8]: an ON state in which the neurons are depolarized and generate spontaneous spikes of membrane potential (known as action potentials), and an OFF state in which the neurons are hyperpolarized and electrically silent (Fig. 1). During SWA, the neurons oscillate in a synchronized pattern between ON and OFF states, generating changes in the surface potential of the cortex that can be monitored using EEG. During the awake state, the firing of cortical neurons is more active and less synchronized than in SWA. One of the key premises of Vyazovskiy and colleagues' study[6], and of prior work[9], is that the OFF state is the hallmark of sleep in cortical neurons.

The authors adopted a straightforward approach to determine what happens to the electrical activity of cortical neurons in sleep-deprived rats: they implanted microwire arrays into the animals' motor and parietal cortices
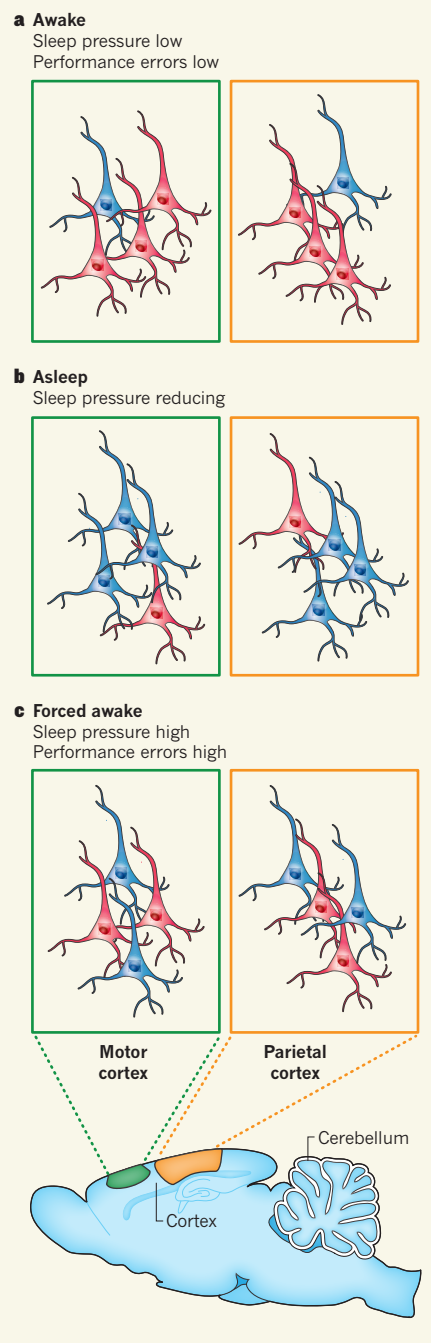
a Awake
Sleep pressure low
Performance errors low

b Asleep
Sleep pressure reducing

c Forced awake
Sleep pressure high
Performance errors high

Motor cortex   Parietal cortex

Cerebellum

Cortex

Figure 1 | Neuronal activity in the rat brain. a, In the awake brain, when the pressure to sleep is low, most neurons in the motor cortex and the parietal cortex are in the ON state (red), as defined by their electrical activity. Only a few are in the OFF electrical state (blue), which is associated with sleep. **b**, In the sleeping brain, the converse is true. **c**, Vyazovskiy et al.[6] report that in awake, sleep-deprived rats, the number of cortical neurons in the OFF state correlates with the pressure to sleep, and that the rats make more errors than fully awake rats in performing a task associated with neurons in the motor cortex. The presence of neurons in the OFF state in the motor cortex did not correlate with the presence of such neurons in the parietal cortex, suggesting that the observed 'switching off' of individual neurons during sleep deprivation is not coordinated across the whole brain.

to measure local field potentials and neural activity during the first four hours of the day. Rats are nocturnal and normally sleep early in the day, so the pressure for sleep steadily increases during the four-hour extension of wakefulness. No method of sleep deprivation in rodents is perfect, but the authors used a previously established approach[10] in which the rats were stimulated by being given new objects to play with, and, importantly, used a detailed video analysis to confirm their success in keeping the animals awake.

Vyazovskiy et al.[6] observed that, in the rats' frontal motor cortex and parietal cortex, the number of field-potential oscillations in the SWA frequency range progressively increased over the sleep-deprivation period, indicating an increase in sleep pressure[11]. Crucially, they also observed that an increasing number of single units — probably single neurons — entered the OFF state (Fig. 1). These findings suggest that, when a rat experiences mounting sleep pressure, individual neurons progressively move into a sleep state even while the animal is awake.

Sleep can be viewed as a global process, one that affects the whole brain. Humans and other mammals adopt a species-specific body position and exhibit reduced sensitivity to external stimuli when asleep, and their brains undergo systemic changes in neural activity patterns, as detected by EEG. Yet sleep also has several localized properties: stimulation of specific areas of the brain can promote or inhibit sleep, and neural activity during sleep shows region-specific patterns.

By taking simultaneous recordings in two areas of the cortex of sleep-deprived rats, Vyazovskiy et al.[6] were able to determine whether OFF periods in those regions were coordinated, as would be expected for global behaviour. Instead, they observed that, although OFF states increased in both regions during sleep deprivation, most occurred in one region but not the other. Even within the same cortical region, some neurons were OFF while others were ON. These data fit our growing understanding of sleep as a local and history-dependent phenomenon[12] — for example, a sleep-like state can exist in certain cortical regions, depending on the prior activity of the neurons in those regions[13,14]. What's more, Vyazovskiy and colleagues' data suggest that the basic unit of sleep is the electrical activity of a single cortical neuron.

The argument that sleep is local matches clinical observations of sleepwalking and other parasomnias in which individuals enter a twilight state between sleep and wakefulness[15]. Nature also provides wonderful examples of animals in which one hemisphere of the brain remains awake while the other moves into SWA[16]— typically birds and marine mammals that need to move continually. These observations also suggest that single neurons can move into a rest state

and that sleep can be initiated locally.

In the final part of their study, Vyazovskiy et al. investigated whether the observed increase in OFF states in the motor cortex of sleep-deprived rats affected the animals' behaviour. This would seem reasonable, given that one of the characteristic features of a sleep-deprived person or animal is a decrease in performance on many tasks. The authors trained the rats to reach for a sugar pellet as a food reward, a task that has previously been shown to increase synaptic strength in the motor cortex[17]. They found that the increasing occurrence of the OFF state correlated with decreased performance in this task. The authors therefore suggest that neurons entering the sleep state in an awake brain are responsible for performance impairments that occur because of sleep deprivation. This is the most speculative conclusion of the study, as it is arguably an intellectual stretch. Nevertheless, the authors' data provide a basis for further testing of this hypothesis. With continuing improvements in optogenetic technologies that allow precise control of targeted cells, it may soon be possible to put restricted groups of neurons to 'sleep' and to determine the impact of this on performance.

To paraphrase Oscar Wilde, scientific truth is rarely pure and never simple. Is it appropriate to think of single neurons as being asleep while the brain is awake? If so, then the physiological mechanisms that govern the ON and OFF states will need a closer look, as will the role of the neuromodulator molecules that switch single neurons between these states. And although it is only anecdotal evidence, I could swear that some of my students can sleep with their eyes wide open! ■

Christopher S. Colwell *is in the Laboratory of Sleep and Circadian Medicine, Department of Psychiatry, School of Medicine, University of California, Los Angeles, Los Angeles, California 90024, USA.*
e-mail: ccolwell@mednet.ucla.edu

1. Altena, E. et al. Prog. Brain Res. **185,** 181–205 (2010).
2. Bloom, H. G. et al. J. Am. Geriatr. Soc. **57,** 761–789 (2009).
3. Zeitzer, J. M., Nishino, S. & Mignot, E. Trends Pharmacol. Sci. **27,** 368–374 (2006).
4. Saper, C. B. et al. Neuron **68,** 1023–1042 (2010).
5. Crocker, A. & Sehgal, A. Genes Dev. **24,** 1220–1235 (2010).
6. Vyazovskiy, V. V. et al. Nature **472,** 443–447 (2011).
7. Steriade, M. Neuroscience **137,** 1087–1106 (2006).
8. McCormick, D. A. Curr. Biol. **15,** R294–R296 (2005).
9. Vyazovskiy, V. V. et al. Neuron **63,** 865–878 (2009).
10. Schiffelholz, T. & Aldenhoff, J. B. Neurosci. Lett. **328,** 41–44 (2002).
11. Dijk, D. J. & Czeisler, C. A. J. Neurosci. **15,** 3526–3538 (1995).
12. Krueger, J. M. et al. Nature Rev. Neurosci. **9,** 910–919 (2008).
13. Rector, D. M., Topchiy, I. & Rojas, M. Sleep **28,** A26 (2005).
14. Hanlon, E. C. et al. Sleep **32,** 719–729 (2009).
15. Mahowald, M. W. & Schenck, C. H. Nature **437,** 1279–1285 (2005).
16. Siegel, J. M. Nature **437,** 1264–1271 (2005).
17. Rioult-Pedotti, M.-S., Friedman, D. & Donoghue, J. P. Science **290,** 533–536 (2000).

REVIEW

On the role of the Agulhas system in ocean circulation and climate

Lisa M. Beal[1], Wilhelmus P. M. De Ruijter[2], Arne Biastoch[3], Rainer Zahn[4] & SCOR/WCRP/IAPSO Working Group 136*

The Atlantic Ocean receives warm, saline water from the Indo–Pacific Ocean through Agulhas leakage around the southern tip of Africa. Recent findings suggest that Agulhas leakage is a crucial component of the climate system and that ongoing increases in leakage under anthropogenic warming could strengthen the Atlantic overturning circulation at a time when warming and accelerated meltwater input in the North Atlantic is predicted to weaken it. Yet in comparison with processes in the North Atlantic, the overall Agulhas system is largely overlooked as a potential climate trigger or feedback mechanism. Detailed modelling experiments—backed by palaeoceanographic and sustained modern observations—are required to establish firmly the role of the Agulhas system in a warming climate.

The greater Agulhas system around southern Africa forms a key component of the global ocean circulation[1–3] (Fig. 1). This system feeds the upper arm of the Atlantic meridional overturning circulation (AMOC) through the leakage of warm, saline waters from the Indian Ocean to the Atlantic[1,4] (Fig. 2). From year to year, Agulhas leakage is dominated by nonlinear, mesoscale dynamics: it is carried by Agulhas rings (formed by an occlusion of the Agulhas Retroflection; Fig. 2), eddies and filaments[5–7]. Over longer periods, theory suggests its variability is associated with the large-scale wind field, in particular with the position of the maximum Southern Hemisphere westerly winds[2,8–10]. These winds are related to the latitude of the oceanic subtropical front (STF), which separates the subtropical gyre from the Antarctic

Circumpolar Current. In essence, if the westerlies shift southwards, as recent data suggest in a warming climate (Fig. 1), then the oceanic 'gateway' between the African continent and the STF expands and leakage from the Indian Ocean to the Atlantic increases[8,9] (Fig. 1 and Box 1). Similarly, a northward shift, as inferred from palaeorecords during glacial periods, would reduce the leakage[11,12].

Model simulations suggest that variability in Agulhas leakage can impact the strength of the Atlantic overturning on a number of time-scales[13–15]. Most significantly, a persistent change in leakage could impact the thermohaline properties of the Atlantic, changing its stratification and its potential for deep convection, and thus altering the AMOC to a new stable state over a period of several hundred years[13,16],
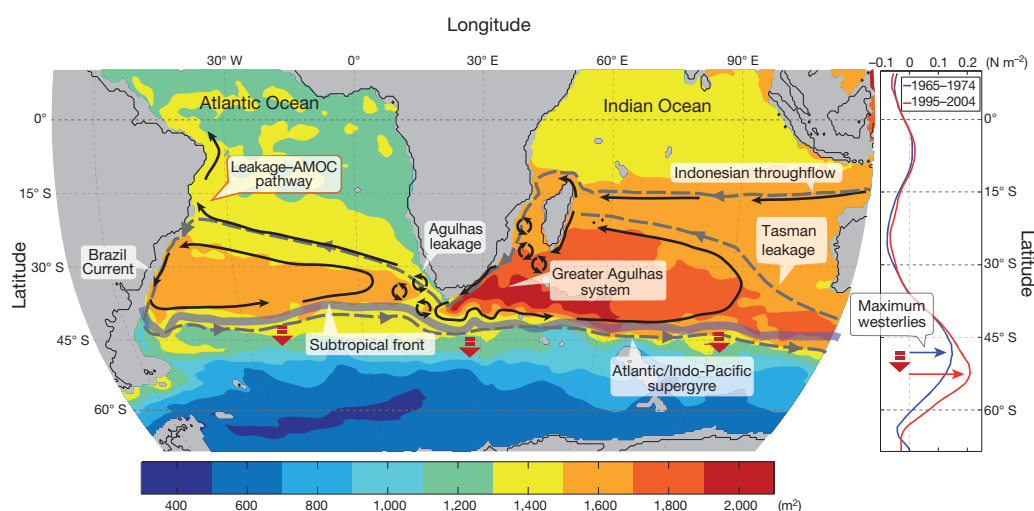

Figure 1 | Agulhas leakage affected by westerly winds and position of subtropical front. Schematic of the greater Agulhas system embedded in the Southern Hemisphere supergyre. Background colours show the mean subtropical gyre circulation, depicted by climatological dynamic height integrated between the surface and 2,000 dbar, from the CARS database[38]. Black arrows and labels illustrate significant features of the flow. An outline of the Southern Hemisphere supergyre is given by the grey dashed line. The plot on the right shows the southward expansion of the Southern Hemisphere westerlies over a 30-yr period, from the CORE2 wind stress[92] averaged between longitudes 20° E and 110° E (Indian Ocean sector). The expected corresponding southward shift of the subtropical front (STF) is illustrated by red dashed arrows and would affect Agulhas leakage (shown as eddies) and the pathway between leakage and the AMOC, which is highlighted with a red box.

[1]Rosenstiel School of Marine and Atmospheric Science, University of Miami, 4600 Rickenbacker Causeway, Florida 33149, USA. [2]Institute for Marine and Atmospheric Research, Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands. [3]Leibniz-Institut für Meereswissenschaften (IFM-GEOMAR), Düsternbrooker Weg 20, 24105 Kiel, Germany. [4]Institució Catalana de Recerca i Estudis Avançats (ICREA), Institut de Ciència i Tecnologia Ambientals (ICTA), Departament de Física, Universitat Autònoma de Barcelona, Bellaterra E-08193, Spain.
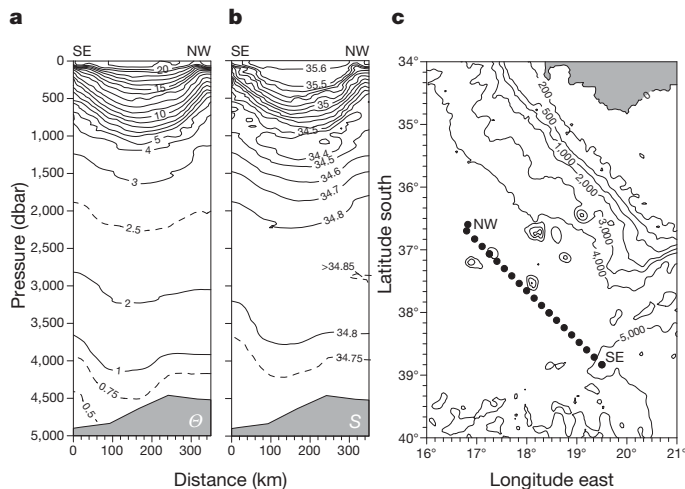*Lists of authors and affiliations appear at the end of the paper.

Figure 2 | Warm, saline Agulhas waters influence temperature and salinity in the Atlantic over the full depth of the water column. a, b, Vertical sections of the potential temperature (Θ; **a**) and salinity (S; **b**) of Agulhas ring Astrid, measured in the Cape basin off South Africa, March 2000[5]. Agulhas waters trapped in the ring displace otherwise flat surfaces of Θ and S by up to 500 m. The sea bed is shaded grey. **c,** Hydrographic station positions are indicated by dots on the chart, where the African continent is shaded grey.

with direct implications for climate. According to these simulations, a stronger Agulhas leakage (saltier Atlantic) sustains a stronger and more stable AMOC.

Fluctuations in the strength of Agulhas leakage over the late-Pleistocene epoch have been inferred through an assemblage of planktonic foraminifera characteristic of modern Agulhas waters in marine sediment records[17]. The reconstructions suggest that there was less Agulhas water leaking into the South Atlantic during glacial periods, when the STF was shifted several degrees northwards of its present-day position[18]. Furthermore, leakage started to increase during late-glacial conditions, several thousand years before the glacial ice volume fully disappeared[11,17,18]. This maximum during glacial terminations suggests that Agulhas leakage may have had a role in the rapid resumption of interglacial climate, presumably through its influence on the AMOC[14,17,19].

There is evidence that Agulhas leakage is increasing under anthropogenic climate change. Satellite and hydrographic data show a southward expansion of the Indian Ocean subtropical gyre[20] and a warming trend in the Agulhas system since the 1960s[10], in correspondence with a southward shift of the STF. There has also been an increase in eddy kinetic energy in the southeast Atlantic since the early 1990s, indicative of more rings and eddies associated with increased leakage[10]. Hindcast simulations suggest that these observed changes may accompany an increase in Agulhas leakage of the order of 1.4–4 Sv per decade[9,10] ($1 \text{ Sv} = 1 \times 10^6 \text{ m}^3 \text{ s}^{-1}$), which could lead to a significant strengthening of the AMOC[13]. This is a profound finding, because it presents a plausible mechanism to stabilize the AMOC at a time when anthropogenic warming and accelerated Greenland ice-sheet melting is predicted to weaken it[21,22].

In this Review, we focus on the past decade of research on the Agulhas system, highlighting its potential global climate impacts in particular. We find observational and palaeoproxy evidence that Agulhas leakage is enhanced in a warming climate and could precipitate rapid increases in the AMOC and atmospheric carbon dioxide. Theory and simulations show how Agulhas leakage is mechanistically linked to the strength of the Indian Ocean subtropical gyre, to changes in Southern Hemisphere winds and to changes in the AMOC. There are large uncertainties in our current understanding and we suggest that sustained observations, as well as climate process studies with high resolution in the Agulhas region, are badly needed. A careful method for simulating changes in leakage (accounting for both temperature and salinity fluxes) is critically important in regards to the response of the AMOC, as is an exploration

BOX 1
Dynamics of Agulhas leakage

As a western boundary current, the Agulhas Current is primarily driven by positive wind stress curl over the subtropical Indian Ocean. However, beyond the tip of Africa, there is no western boundary to steer the current southwards, towards the latitude where the westerlies are a maximum and the wind stress curl is zero (Fig. 1). Hence, without inertia the full Agulhas Current would turn west into the Atlantic and feed the Southern Hemisphere supergyre[8]. Conversely, for large inertia, or with the latitude of zero wind stress curl much closer to the continent, the full Agulhas Current would loop back into the Indian Ocean. In this case, there is no leakage, owing to 'inertial choking'. Lower inertia would lead to partial leakage. Hence, the main controls on retroflection and leakage are the latitude of maximum westerlies and the southward inertia of the Agulhas Current at separation, which are both largely determined by the strength and position of the wind field over the Indian Ocean. Instabilities, interaction with bathymetry and other mesoscale, nonlinear dynamics are also important, generating Agulhas rings that propagate westwards, conveying most of the leakage into the Atlantic[2,95,96].

When interpreting glacial–interglacial cycles in the Agulhas system implied from palaeoproxy data, or present-day changes implied from model simulations, the combination of these controls should be taken into account. For instance, if the wind pattern is fixed but its strength reduced, then a weaker Agulhas Current would result in more leakage, owing to the smaller inertial overshoot[97,98]. Similarly, for increased wind strength a stronger Agulhas Current combines with less leakage. But if the wind pattern shifts northwards and weakens (for example in relation to glacial climate[99]) then the Agulhas Current would be weaker at separation, and the latitude of zero wind stress curl would simultaneously be closer to Africa. The weaker Agulhas Current may then still have enough inertia to reach the closer wind curl change, so that a weaker Agulhas Current combines with less leakage[17,91]. In present-day climate change, global wind data point to a southward expansion of the wind pattern (a southward shift of the westerlies) but no robust trend in wind curl over the Indian Ocean. In such a situation, the latitude of zero wind stress curl or maximum westerlies is farther south and leakage increases, but the response of the Agulhas Current and its change in inertia is unclear[9,10].

To simulate Agulhas leakage correctly, it is essential that inertial mechanisms and ring formation—that is, nonlinear dynamics—are resolved. However, most Intergovernmental Panel on Climate Change (IPCC)-type simulations have coarse resolution (large eddy viscosity). Agulhas retroflection appears, but as a result of 'viscous choking' instead of inertial choking. As a result, the amplitudes of Agulhas transport and leakage and their sensitivity to change may be completely wrong[100]. An effective resolution of at least 0.1° is necessary to resolve Agulhas leakage and could be achieved by a range of regional and global numerical grids[15,70].

into the interplay of Southern Hemisphere winds and Agulhas leakage in forcing a double response of the ocean in a changing climate.

The Agulhas system and ocean circulation

The Agulhas Current is the western boundary current of the southern Indian Ocean subtropical gyre (Fig. 1), and is primarily driven by the large-scale pattern of wind stress curl between the southeast trade winds and the Southern Hemisphere westerlies[3]. The STF at a latitude of about 45° S demarks the southern extent of the gyre and lies well south of the tip of the African continental shelf at 37° S. The Agulhas Current flows to the southwest along the east coast of South Africa as a narrow, fast boundary current until it separates from the continent, looping anti-clockwise south of Africa and feeding back into the Indian Ocean as the eastward Agulhas Return Current (Box 1). This loop, known as the

Agulhas Retroflection[23], sheds rings, eddies and filaments to the west, representing a leakage of Indo-Pacific waters into the Atlantic down to depths of more than 2,000 m (refs 1, 5–7; Figs 2 and 3).

The Agulhas Current is fed in the main by recirculating subtropical gyre waters, but also by waters from the Red and Arabian seas, from the Indonesian throughflow and from the equatorial Indian Ocean via Mozambique Channel eddies and the East Madagascar Current (EMC)[24,25] (Fig. 3). Long-term moorings in the narrows of the Mozambique Channel show that four or five large (350-km) anticyclonic eddies drift southwards through the channel per year, carrying a mean transport of 17 Sv (ref. 26). Interannual variability is high (9 Sv), owing to upstream variations related to the phase of the Indian Ocean Dipole mode, an intrinsic mode of coupled climate variability in the tropical Indian Ocean[26]. The EMC is less well measured. At 20° S there is an estimated transport of 20 Sv (refs 27, 28), whereas close to the southern tip of Madagascar the geostrophic transport is increased to about 35 Sv (ref. 29). The Agulhas Current itself has a mean transport of 70 Sv at 32° S (ref. 30). Its variability on seasonal to longer timescales remains unknown. Estimates of Agulhas leakage are highly uncertain, ranging between 2 and 15 Sv, with about four to six Agulhas rings shed per year[1,2,31].

There has been conflicting evidence over the years for the existence of an EMC retroflection[32–36]. Dynamically, the large-scale wind field does not require the EMC to feed an eastward interior flow as it does at the latitude of the Agulhas Retroflection. Most recent research points to the EMC linking to the Agulhas Current by means of westwards-drifting

eddies and dipoles shed off the tip of Madagascar[34,35]. Although there is a shallow (<200-m) eastward flow (the South Indian Countercurrent; Fig. 3) around 25° S (ref. 37) that may connect intermittently with the EMC[36], mean dynamic topography, satellite and water mass measurements show it to be part of a basin-wide recirculation to the south and east of Madagascar that is fed by waters from the southwest[29,37–39].

Observations have recently shown, as predicted dynamically[8], that the Agulhas system and the southern Indian Ocean subtropical gyre are embedded in a Southern Hemisphere supergyre that connects the Atlantic, Indian and Pacific basins[31,38,40] (Fig. 1). The Atlantic and Indian oceans are connected south of Africa by westward Agulhas leakage and the eastward South Atlantic Current. Hence, a proportion of Agulhas leakage remains in the South Atlantic subtropical gyre, some ends up in the supergyre, circulating back into the Indo-Pacific via the South Atlantic Current, and the remainder feeds into the surface arm of the AMOC and crosses the equator.

The strength of the connectivity between Agulhas leakage and the AMOC is important for global climate impacts but is beyond direct observations. As a source of the AMOC, Agulhas leakage (warm-water route) competes with intermediate waters from the Drake Passage (cold-water route)[41,42]. Several hydrographic inverse models have implied that the cold-water route dominates[43], although some previous hydrographic studies concluded otherwise[41]. A careful model-data reanalysis[4] suggests that the AMOC is largely fed by Agulhas leakage and that inverse models deal poorly with large, opposing flows such as those found south of Africa (Figs 1 and 3). This is because hydrographic inverse models work
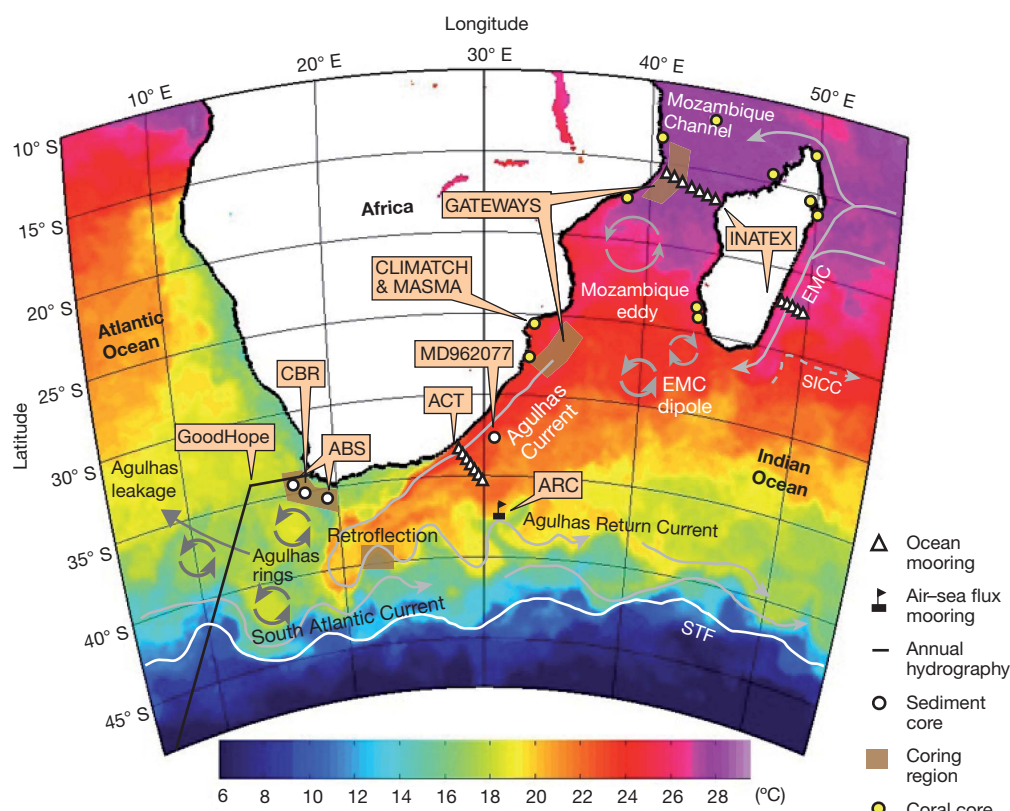

Figure 3 | Agulhas leakage advects warm, saline waters into the Atlantic, predominantly through Agulhas rings. Sea surface temperature (SST) for 23 May 2009, showing water at 23–25 °C in the Agulhas Current and Retroflection. North–south migrations of the STF (related to wind curl changes) and/or the retroflection loop (related to varying Agulhas strength) can choke or open the "leakage gap" between Africa and the STF. Leakage takes place largely by means of Agulhas rings. Main circulation features and observation programs are highlighted. Monitoring and repeat measurement programmes in the region include Agulhas Current Time-series (ACT), Indian–Atlantic Exchange in Present and Past Climate (INATEX), the Agulhas Return Current (ARC)

air–sea flux buoy and the GoodHope repeat-hydrography line. Palaeoceanographic measurements are being collected in four (coring) regions under the European GATEWAYS programme. Individual sediment core locations are marked for the Cape Basin record[17] (CBR), MD962077[12] and Agulhas Bank splice[90] (ABS). Coral cores are being collected under the Climate and Anthropological Change (CLIMATCH) and Marine and Coastal Science for Management (MASMA) programmes. SICC is the shallow South Indian Countercurrent. SST data are from the NAVOCEANO K10 analysis with combined satellite infrared and microwave measurements, made available through the GHRSST project[93]. EMC, East Madagascar Current.

to balance net section-wide fluxes (and are underdetermined); hence, Agulhas leakage is obscured by the large throughflow of the Antarctic Circumpolar Current. Maps of mean dynamic height from the Argo database also support a direct pathway for Agulhas leakage from the Cape basin to the equator (Fig. 1).

Mechanistic linkages to climate

The mechanisms by which the Agulhas Current and its leakage may be linked to AMOC and global climate have been explored in several ocean and coupled climate models. Variability in Agulhas leakage can impact the strength of the AMOC in a number of different ways: through buoyancy forcing, associated wind stress changes and planetary-wave perturbations.

The depression of density surfaces associated with anticyclonic Agulhas rings entering the Atlantic basin instigates the propagation of planetary waves that perturb the AMOC[15,44]. Such perturbations can contribute to short-term variability in the AMOC, as might be seen in the time series at 26° N for instance[45], but are also integrated into interannual-to-decadal oscillations in the AMOC[15], perhaps because Agulhas ring shedding is modulated by Indian Ocean Dipole modes and El Niño/Southern Oscillation modes[34,46].

Simulations show that buoyancy forcing associated with warm and saline Agulhas leakage waters acts to strengthen the AMOC by enhancing the Atlantic meridional pressure gradient and preconditioning the North Atlantic for deep convection[13,16]. Although Agulhas heat and salt anomalies are largely density compensated as they enter the South Atlantic, progressive atmospheric heat loss creates an increasing positive density anomaly as salt is left behind and waters are advected northwards[13,47]. In this way, an increase in leakage initiates an increase in overturning strength 15–30 yr later, the advective timescale for the excess salt to reach the North Atlantic[13,47]. It takes another several hundred years for the AMOC to adjust fully to a new stable state, as buoyancy fluxes and (irreversible) mixing rates re-establish a balance[13,48].

It is difficult to separate the impact of wind forcing on the AMOC from the buoyancy effect of Agulhas leakage, because model studies show that the strength and position of the Southern Hemisphere wind systems seem to affect the strengths of both Agulhas leakage and the AMOC directly[2,8,9,49,50]. For example, in a warming climate the Southern Hemisphere westerlies would tend to shift polewards and the AMOC will have a double response: (1) an expansion of the gateway between Africa and the STF allows increased Agulhas leakage that, through buoyancy forcing, increases the AMOC[13]; and (2) increased northward surface-layer Ekman transport in the Southern Ocean increases upwelling that theoretically pulls directly on the southward deep arm of the AMOC (Drake Passage effect[49]). Climate model experiments of coarse resolution, where the wind systems are shifted to emulate glacial–interglacial cycles or projected anthropogenic warming, suggest that a warmer climate is associated with a stronger AMOC (and a saltier Atlantic) and that a cooler climate is associated with a weaker AMOC, presumably through an unknown combination of mechanisms (1) and (2) and others (for example sea ice and air–sea fluxes) that are unrelated to the Agulhas system[50,51].

Changes in Southern Ocean winds and buoyancy fluxes are also thought to regulate the ocean's ability to store carbon[52,53], through their impact on deep upwelling (ventilation) and the AMOC. Hence, Agulhas leakage could influence atmospheric carbon dioxide concentrations through its link to the AMOC, or through a direct impact on the density distribution in the sub-Antarctic region. So far, simulations have produced conflicting results regarding the relative importance of wind and buoyancy in governing atmospheric carbon dioxide[54]. The broader picture is further complicated by the role of eddies, which are unresolved in climate models but can carry significant buoyancy fluxes and are clearly important in controlling Agulhas leakage. In the Southern Ocean, eddies can short-circuit the link between Ekman transport, deep upwelling and the AMOC, thereby significantly reducing the sensitivity of the AMOC and carbon dioxide to winds[55,56]. This means that climate models overestimate the control winds have over the AMOC and carbon dioxide variability[57]; hence, Agulhas

leakage and buoyancy fluxes may have a more important role than presently recognized.

Observed interactions with climate

Western boundary current systems support the highest air–sea fluxes in the world, and the Agulhas system is no exception. Latent and sensible heat fluxes increase three to five times over the SST fronts associated with the Agulhas Current, Retroflection and Return Current[58]. There is a clear impact on surface winds and clouds, with higher wind stress, a deeper marine–atmospheric boundary layer and convective clouds over the warmer waters[59,60]. As a result of enhanced atmospheric baroclinicity, the Agulhas system influences storm development, storm tracks and the regional atmospheric circulation[61,62]. It can also contribute to extreme rainfall events and tornadoes over southern Africa[63]. Increased rainfall events are correlated with warm anomalies in the Agulhas system associated with subtropical Indian Ocean Dipole and El Niño/Southern Oscillation cycles[64,65]. A 300-yr coral record suggests that overall warming of the Agulhas system since the 1970s may have increased the sensitivity of the African hydrological cycle to El Niño/Southern Oscillation[66].

Agulhas leakage may also be modulated by interannual modes of variability in the Indian and Pacific oceans. The Indian Ocean tropical and subtropical gyres shift in latitude and strengthen during a negative Indian Ocean Dipole phase (La Niña) and weaken during a positive phase (El Niño) through planetary-wave teleconnection processes[46,67], as observed in satellite data. These gyre changes strengthen or weaken the sources of the Agulhas Current, affecting the frequency of Mozambique Channel and EMC eddies (anticyclonic anomalies) that can propagate downstream into the retroflection, where they tend to trigger Agulhas rings[34,46]. Along the Agulhas Current, these anomalies manifest as strongly barotropic solitary meanders that propagate downstream with a mirrored cyclone on the inshore side[30,68–70] and destabilize the retroflection. In model simulations, the resulting fluctuations in Agulhas leakage cause decadal-scale variability in the AMOC[15].

Observations of interactions between the Agulhas system and climate over longer periods are based on palaeoceanographic records that currently span the past 570,000 yr (Fig. 4). Systematic changes in the abundance of planktonic foraminiferal species associated with modern-day Agulhas waters are recorded in marine sediment cores positioned along the Agulhas ring corridor and used as a proxy for leakage strength[17,71] (Agulhas fauna). Leakage events can be linked to glacial–interglacial cycles using oxygen isotope stratigraphy and to variations in deep-ocean ventilation using carbon isotope ratios and opal fluxes[12,17,72,73]. Variations in the ratio of subtropical to subpolar planktonic species taken from sites close to the modern-day STF can be used as a proxy for its latitudinal position in the past. Together, these palaeo-proxies point to three conclusions: Agulhas leakage peaks at glacial terminations, changes in leakage are accompanied by latitudinal migrations of the STF by several degrees, and changes in leakage seem to be accompanied by changes in deep-ocean ventilation, suggesting that the AMOC strengthens during glacial–interglacial transitions[72,73]. Enhanced ventilation through Southern Ocean upwelling has also been correlated with the deglacial rise in concentrations of atmospheric carbon dioxide[73]. Importantly, peaks in Agulhas leakage appear during deglaciation, while ice volume is still substantial, suggesting that leakage may have had a role in triggering the associated resumption of the AMOC and the relatively rapid rise in carbon dioxide[17]. During the last deglaciation, the delay in and then abrupt warming of the North Atlantic (transition from Heinrich event 1 to the Bølling warm event) has been attributed to changes in the AMOC[74], and a high-resolution sediment core[19] supports model studies[13,14] that point to Agulhas leakage as the trigger. Furthermore, during two of the most severe glacials 340,000 and 420,000 yr ago (MISs 10 and 12), the STF seemed to have shifted northwards by about 7°, to reach its lowest latitude in the past 800,000 yr, and Agulhas fauna suggests that leakage was weaker than ever and possibly ceased[12] (Fig. 4). This implies that Agulhas leakage may have a role in modulating the severity of glacial periods through its influence on the AMOC[7,75].
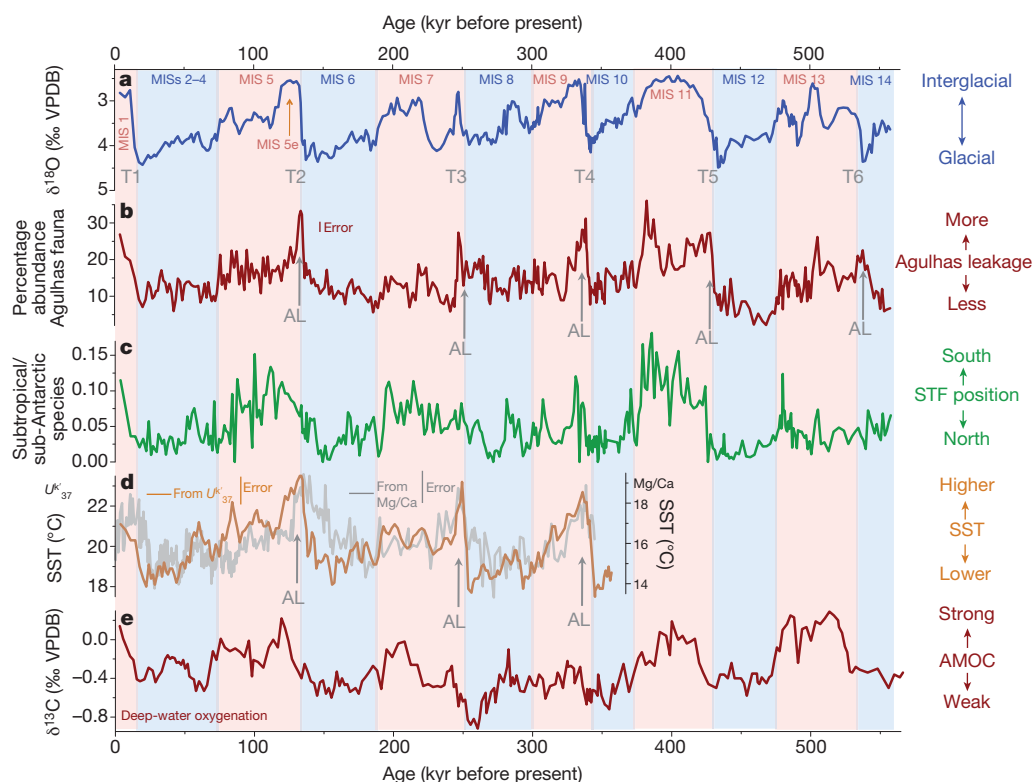
Figure 4 | Palaeoceanographic time series from the Agulhas leakage corridor spanning the last 570,000 yr. a, Stable oxygen isotope profile ($\delta^{18}O = (^{18}O/^{16}O)_{sample}/(^{18}O/^{16}O)_{VPDB} - 1$; Vienna PeeDee Belemnite (VPDB) standard) of benthic foraminifera from the CBR[17] represents a proxy for glacial–interglacial variations in global climate (highlighted by vertical blue/red shading). Marine Isotope Stages (MISs) are labelled. T1–T6 mark terminations of the past six glacial periods (MISs 2, 6, 8, 10, 12 and 14). **b**, Abundance of tropical planktonic foraminiferal marker species along the CBR[17] indicate maximum Agulhas leakage (AL) during glacial terminations T2–T6. Standard error (1.73%) is illustrated. **c**, Ratio of subtropical to sub-Antarctic planktonic marker species along the CBR[17], thought to be related to north–south migrations of the STF. **d**, SST derived from temperature-sensitive biomarkers ($U^{k'}_{37}$) produced by prymnesiophyceae algae along the CBR[17] (brown line), and Mg/Ca ratios in planktonic foraminifera from the ABS record[90] (grey line). Both reconstructions show maximum SST during glacial terminations, coinciding with Agulhas leakage events. SST reconstructions diverge during glacial periods, possibly corresponding to changes in seasonality. Error bars represent a combination of analytical and calibration errors. **e**, Record of benthic $\delta^{13}C$ (($^{13}C/^{12}C)_{sample}/(^{13}C/^{12}C)_{VPDB} - 1$) from ODP Site 846 in the deep Pacific[94] is thought to be linked to ventilation of the deep Pacific[72], which is related to the strength of the AMOC. With this interpretation, AMOC strength seems to increase at each glacial termination, leading to the hypothesis that Agulhas leakage may stimulate the AMOC. All records are synchronized using standard oxygen isotope stratigraphy.

There is recent evidence pointing to a trend of increasing Agulhas leakage[9,10] concomitant with a polewards shift in the westerly winds over the past three decades, which is projected to continue over the twenty-first century under anthropogenic forcing[51,76]. This evidence is based on satellite data, far-field hydrographic data and hindcast simulations (ocean models with prescribed atmospheric forcing over the past 50 yr), because no modern *in situ* observations of Agulhas leakage exist. The Agulhas system seems to have been warming since the 1960s[10,20], and increased eddy kinetic energy in the southeast Atlantic, together with saltier water in the North Brazil Current and crossing the equator, is suggestive of increased leakage[9].

Facing uncertainties

We have seen that Agulhas leakage may have an important role in climate transitions, but it cannot act in isolation from other processes in the climate system. Models suggest that the AMOC also adjusts to the shifting of the Southern Hemisphere westerlies[50] and to Northern Hemisphere ice melt[14,77]. Southern Ocean sea-ice extent and carbon dioxide outgassing are further likely feedbacks[52,78]. The relative importance and timing of these effects, and the interactions among them, are largely unknown. Climate models tend to overestimate wind effects because of unresolved eddies[56,57], and to underestimate surface buoyancy fluxes as a result of parameterizations that suppress deep convection[48]. Moreover, most do not resolve the nonlinear dynamics that control Agulhas leakage (Box 1).

There is ongoing debate about the relative roles of surface buoyancy forcing and wind (and tide) forcing on the AMOC. From energy considerations, it has been argued that buoyancy fluxes have only some kind of organizational role[79,80] by setting water mass stratification, and that wind and tides drive the AMOC by providing the energy needed for the (irreversible, diapycnal) mixing that brings deep water back to the surface. The implication of this conceptual model is that Agulhas leakage is merely a passive respondent, and that AMOC changes are forced elsewhere by changing winds. However, a new study[48] shows that the rate at which energy is transferred within the system is more important than the quantity of available energy, and that the background stratification essentially dictates these rates. In other words, wind and tides represent the fuel of the system and stratification represents the efficiency, such that the rate at which mixing can redistribute mass is key, not the total amount of mixing per se. Hence, in a steady state, the redistribution of mass by surface buoyancy fluxes is balanced by the redistribution of mass by mechanical mixing. In this case, buoyancy fluxes are important and we might expect Agulhas leakage and wind changes to force a double response of the AMOC, as discussed earlier[13,50]. Each acts to strengthen the AMOC in a warming climate and weaken it in a cooling climate, which perhaps explains why the Southern Hemisphere leads in glacial–interglacial transitions. In contrast, Northern Hemisphere ice melt acts in the opposite sense, reducing the AMOC in response to warming. Overall, Southern Hemisphere forcing may dominate the glacial cycle, causing the AMOC to increase during deglaciation[14,72].

Simulations show that the strength and pathway of warm, saline Agulhas leakage through the South Atlantic profoundly affects the sensitivity of the AMOC to leakage changes. Like other subtropical gyres, the South Atlantic gyre has a poleward salt flux via the Brazil Current (reflecting an excess in evaporation over precipitation over the subtropics), but the superposed AMOC has an equatorward salt flux, representing Agulhas leakage waters at the surface and a return of fresher North Atlantic Deep Water below[4,81,82]. Several model studies demonstrate that the AMOC is bistable if it imports its own salt, because this 'salt–advection feedback' allows a collapsed AMOC (as well as an active one) as a stable state[83–85]—although there is no evidence for a completely collapsed AMOC in the palaeorecord[12]. In all these models, the South Atlantic gyre exports salt from the Atlantic basin and the overturning component of the circulation imports a little less salt, such that overall the Atlantic Ocean is evaporative, consistent with air–sea flux climatologies[86]. However, hydrographic data point to a small import of salt (AMOC dominant), inconsistent with an evaporative basin[81,82], whereas most IPCC climate models show AMOC salt export, inconsistent with both the oceanic data and salt–advection feedback[87]. These inconsistencies could represent a serious flaw in predictions of the AMOC response to changes in Agulhas leakage. Theoretically, in a monostable state, if the AMOC shuts down it will eventually recover because it is not stable in a collapsed state. In a bistable state, a shutdown can be permanent, leading us to an altogether different climate state. The lack of salt–advection feedback in IPCC models (and by inference an incorrect Agulhas leakage) could be the reason why these models do not show multiple stable states of the AMOC[87] and may render their predictive capabilities inadequate.

Wind-shift experiments in coarse-resolution climate models seem to suggest that the AMOC has an asymmetric response under present-day climate, such that a northward shift in the Southern Hemisphere westerlies decreases the AMOC by several sverdrups whereas a southward shift seems to have little effect[50]. An explanation could be that the larger Agulhas leakage for a southward wind shift mostly recirculates in the supergyre and does not force the AMOC in the model. The response of the AMOC to changes in leakage is profoundly shaped by the evolution of leakage waters, their associated density anomaly and whether they are transported northwards in the Atlantic[13,81]. It is important that both temperature and salinity anomalies (and their vertical distributions; Fig. 2) are simulated accurately, and with realistic air–sea fluxes, to correctly predict the sensitivity of the AMOC to leakage. For instance, introducing only surface salinity (a positive density anomaly) into the South Atlantic as a proxy for Agulhas leakage initially destabilizes the meridional pressure gradient and slows (or even halts) the AMOC, an effect not present when appropriate temperature–salinity anomalies are used[13,16,88]. For the same reasons, oceanic turbulent buoyancy fluxes initiated by eddy–eddy and eddy–mean flow interactions in the leakage region are important, as is the mean path of the leakage. So far, climate models do not accurately resolve the processes controlling leakage and its pathways into the supergyre and the AMOC.

The palaeorecord likewise has inconsistencies, such as the indication of an anomalous SST increase in the southeast Atlantic during one of the superglacials[89] (MIS 10), alluding to the advection of subtropical waters well into the sub-Antarctic zone during a period when Agulhas fauna suggests leakage was at a minimum. A core at the northern edge of the Agulhas ring corridor suggests that leakage may have increased much earlier than during glacial terminations[90]. Radiogenic isotopes that serve as a fingerprint in sea-floor sediments of overlying Agulhas water suggest that the position of the Agulhas Retroflection at the Last Glacial Maximum was the same as today[91]. Despite these inconsistencies in detail, the palaeorecord (Fig. 4) provides the strongest evidence yet for a connection between Agulhas leakage, the AMOC and changing climate.

Outlook

Many of the uncertainties in our current understanding of the linkages between the Agulhas system and climate can be addressed with carefully

placed observations and well-designed numerical experiments. Observational programmes to capture long-term transport of the Agulhas Current and air–sea fluxes over the Agulhas Return Current began in 2010 (Fig. 3), and extensive hydrographic observations over the greater Agulhas system have been obtained over the past three years under the auspices of the Agulhas and Somali Current Large Marine Ecosystems Project. However, although there are annual hydrographic occupations across the Cape basin (French/Russian/South African GoodHope programme), direct continuous measurements of Agulhas leakage are very difficult because of its intermittent nature and variable location. Monitoring of the leakage embedded in the South Atlantic close to 25° S seems appropriate, as this is far enough north that difficult-to-capture eddy fluxes are small and the AMOC signal (and salt flux) may be clearer above a weaker gyre signal (weak Brazil Current). The international South Atlantic Meridional Overturning Circulation group is developing a proposal for a basin-wide AMOC array for the South Atlantic, similar to that at 26° N.

Long-term in situ observations of the Agulhas Current and its sources in the Mozambique Channel and EMC are underway through the US ACT and Dutch INATEX programmes (Fig. 3). In support of these observations, closer explorations of the link between Indian Ocean variability, Agulhas Current and Agulhas leakage under different simulated wind regimes are needed. Ongoing efforts to collect palaeoproxy data (European GATEWAYS Programme) aim to cover the trajectory of the Agulhas Current and its sources to be able to discern between local dynamical effects, large-scale wind effects and upstream variability (Fig. 3). Coral cores from the Dutch CLIMATCH and WIOMSA MASMA programmes will provide information about changes in upstream ocean temperature and the hydrological cycle on centennial timescales. A range of numerical simulations of (resolved) leakage under warming conditions (and shifting winds), also beginning with both glacial and interglacial climate conditions, could further help us understand the links among Agulhas Current, Agulhas leakage and the AMOC and its dependence on the background climate state.

Numerical experiments that explore the sensitivity of the AMOC to Agulhas leakage (with realistic temperature–salinity distributions) in the absence of Southern Hemisphere wind effects are also needed. This might be done by changing the easterly winds over the Indian Ocean or by prescribing leakage, although it is not clear how the latter might be achieved. High resolution over the Agulhas region is a necessity. In terms of palaeorecords, resolution on millennial timescales and higher is needed to examine the sequence of changes in warming, leakage, AMOC, atmospheric carbon dioxide and ice volume, to try to unravel cause and effect.

To identify potential drivers of climate change, we need to understand better the relative strengths and interactions between AMOC forcing in the Northern Hemisphere and that in the Southern Hemisphere. Current efforts largely concentrate on the North Atlantic as the presumed driver of the AMOC. It is time to take a more holistic approach and in particular consider the Southern Ocean and Agulhas leakage as potentially powerful modulators of the AMOC and climate.

1. Gordon, A. L., Weiss, R. F., Smethie, W. M. & Warner, M. J. Thermocline and intermediate water communication between the South Atlantic and Indian Oceans. *J. Geophys. Res.* **97,** 7223–7240 (1992).
2. De Ruijter, W. P. M. *et al.* Indian-Atlantic inter-ocean exchange: dynamics, estimation, and impact. *J. Geophys. Res.* **104,** 20885–20910 (1999). **This work reviews the dynamics and variability of the Indian–Atlantic connection and sets the stage for further research into its wider impacts.**
3. Lutjeharms, J. R. E. *The Agulhas Current* (Springer, 2006).
4. Donners, J. & Drijfhout, S. S. The Lagrangian view of South Atlantic interocean exchange in a global ocean model compared with inverse model results. *J. Phys. Oceanogr.* **34,** 1019–1035 (2004).
5. Van Aken, H. M. *et al.* Observations of a young Agulhas ring, Astrid, during MARE in March 2000. *Deep-Sea Res. II* **50,** 167–195 (2003).
6. Schouten, M. W., De Ruijter, W. P. M., van Leeuwen, P. J. & Lutjeharms, J. R. E. Translation, decay, and splitting of Agulhas Rings in the south-eastern Atlantic Ocean. *J. Geophys. Res.* **105,** 21913–21925 (2000).

7. Boebel, O. *et al.* The Cape Cauldron, a regime of turbulent inter-ocean exchange. *Deep-Sea Res. II* **50,** 57–86 (2003).
   This work describes the mesoscale turbulent regime in the Cape basin for the first time, where important eddy mixing and air–sea interaction transform Agulhas leakage waters into those that take part in the AMOC.

8. De Ruijter, W. P. M. Asymptotic analysis of the Agulhas and Brazil current systems. *J. Phys. Oceanogr.* **12,** 361–373 (1982).

9. Biastoch, A., Böning, C. W., Lutjeharms, J. R. E. & Schwarzkopf, F. U. Increase in Agulhas leakage due to pole-ward shift of the Southern Hemisphere westerlies. *Nature* **462,** 495–498 (2009).
   This was the first global hindcast simulation, with an eddy-resolving Agulhas nest, to show that Agulhas leakage is likely to have increased under anthropogenic forcing.

10. Rouault, M., Penven, P. & Pohl, B. Warming in the Agulhas Current system since the 1980's. *Geophys. Res. Lett.* **36,** L12602 (2009).
    In this work, satellite observations and a regional model are used to show that the Agulhas system has been warming, consistent with increased Agulhas Current transport.

11. Berger, W. H. & Wefer, G. in *The South Atlantic: Present and Past Circulation* (eds Wefer, G., Berger, W. H. & Webb, D. J.) 363–410 (Springer, 1996).

12. Bard, E. & Rickaby, R. E. M. Migration of the subtropical front as a modulator of glacial climate. *Nature* **460,** 380–383 (2009).

13. Weijer, W., De Ruijter, W. P. M., Sterl, A. & Drijfhout, S. S. Response of the Atlantic overturning circulation to South Atlantic sources of buoyancy. *Global Planet. Change* **34,** 293–311 (2002).
    This work is the only rigorous exploration of the effect of Agulhas waters on the stability of the AMOC (in the absence of changed wind forcing).

14. Knorr, G. & Lohmann, G. Southern Ocean origin for the resumption of Atlantic thermohaline circulation during deglaciation. *Nature* **424,** 532–536 (2003).
    Using an ocean model, the authors of this work find that abrupt resumption of the interglacial AMOC mode is triggered by increased mass transport from the Pacific and Indian (Agulhas leakage) oceans into the South Atlantic.

15. Biastoch, A., Böning, C. W. & Lutjeharms, J. R. E. Agulhas Leakage dynamics affects decadal variability in Atlantic overturning circulation. *Nature* **456,** 489–492 (2008).

16. Weijer, W., De Ruijter, W. P. M. & Dijkstra, H. A. Stability of the Atlantic overturning circulation: competition between Bering Strait freshwater flux and Agulhas heat and salt sources. *J. Phys. Oceanogr.* **31,** 2385–2402 (2001).

17. Peeters, F. J. C. *et al.* Vigorous exchange between the Indian and Atlantic oceans at the end of the past five glacial periods. *Nature* **430,** 661–665 (2004).
    This work links glacial terminations with increased Agulhas leakage using an assemblage of subtropical foraminifera to trace changes in leakage over the past 550,000 yr.

18. Flores, J. A., Gersonde, R. & Sierro, F. J. Pleistocene fluctuations in the Agulhas Current Retroflection based on the calcareous plankton record. *Mar. Micropaleontol.* **37,** 1–22 (1999).

19. Chiessi, C. M. *et al.* South Atlantic interocean exchange as the trigger for the Bølling warm event. *Geology* **36,** 919–922 (2008).

20. Alory, G., Wijffels, S. & Meyers, G. Observed temperature trends in the Indian Ocean over 1960–1999 and associated mechanisms. *Geophys. Res. Lett.* **34,** L02606 (2007).

21. Gregory, J. M. *et al.* A model intercomparison of changes in the Atlantic thermohaline circulation in response to increasing atmospheric CO_2 concentration. *Geophys. Res. Lett.* **32,** L12703 (2005).

22. Stammer, D. Response of the global ocean to Greenland and Antarctic ice melting. *J. Geophys. Res.* **113,** C06022 (2008).

23. Gordon, A. L. The brawniest retroflection. *Nature* **421,** 904–905 (2003).

24. Song, Q., Gordon, A. L. & Visbeck, M. Spreading of the Indonesian Throughflow in the Indian Ocean. *J. Phys. Oceanogr.* **34,** 772–792 (2004).

25. Beal, L. M., Chereskin, T. K., Lenn, Y. D. & Elipot, S. The sources and mixing characteristics of the Agulhas Current. *J. Phys. Oceanogr.* **36,** 2060–2074 (2006).

26. Ridderinkhof, H. *et al.* Seasonal and interannual variability in the Mozambique Channel from moored current observations. *J. Geophys. Res.* **115,** C06010 (2010).

27. Swallow, J., Fieux, M. & Schott, F. The boundary currents east and north of Madagascar 1: geostrophic currents and transports. *J. Geophys. Res.* **93,** 4951–4962 (1988).

28. Donohue, K. A. & Toole, J. M. A near-synoptic survey of the Southwest Indian Ocean. *Deep-Sea Res. II* **50,** 1893–1931 (2003).

29. Nauw, J. J., van Aken, H. M., Webb, A., Lutjeharms, J. R. E. & De Ruijter, W. P. M. Observations of the southern East Madagascar Current and undercurrent and countercurrent system. *J. Geophys. Res.* **113,** C08006 (2008).

30. Bryden, H. L., Beal, L. M. & Duncan, L. M. Structure and transport of the Agulhas Current and its temporal variability. *J. Oceanogr.* **61,** 479–492 (2005).
    In this work, the first time series of Agulhas Current transport finds a dominant meander mode at 50–70 d.

31. Richardson, P. L. Agulhas Leakage into the Atlantic estimated with subsurface floats and surface drifters. *Deep-Sea Res. I* **54,** 1361–1389 (2007).

32. Lutjeharms, J. R. E. Remote sensing corroboration of retroflection of the East Madagascar Current. *Deep-Sea Res. I* **35,** 2045–2050 (1988).

33. Quartly, G. D. & Srokosz, M. A. SST observations of the Agulhas and East Madagascar retroflections by the TRMM microwave imager. *J. Phys. Oceanogr.* **32,** 1585–1592 (2002).

34. De Ruijter, W. P. M. *et al.* Eddies and dipoles around South Madagascar: formation, pathways and large-scale impact. *Deep-Sea Res. I* **51,** 383–400 (2004).
    This work shows that both anticyclones and cyclones are commonly shed around Madagascar and propagate into the Agulhas system.

35. Quartly, G. D., Buck, J. J. H., Srokosz, M. A. & Coward, A. C. Eddies around Madagascar: the retroflection reconsidered. *J. Mar. Syst.* **63,** 115–129 (2006).

36. Siedler, G. *et al.* Modes of the southern extension of the East Madagascar Current. *J. Geophys. Res.* **114,** C01005 (2009).

37. Palastanga, V., van Leeuwen, P. J., Schouten, M. W. & De Ruijter, W. P. M. Flow structure and variability in the subtropical Indian Ocean: instability of the South Indian Ocean Countercurrent. *J. Geophys. Res.* **112,** C01001 (2007).

38. Ridgway, K. R. & Dunn, J. R. Observational evidence for a Southern Hemisphere oceanic supergyre. *Geophys. Res. Lett.* **34,** L13612 (2007).
    This work reports observation of the Atlantic/Indo-Pacific supergyre in which a large proportion of Agulhas leakage takes part.

39. Palastanga, V., Dijkstra, H. A. & De Ruijter, W. P. M. Inertially induced connections between subgyres in the South Indian Ocean. *J. Phys. Oceanogr.* **39,** 465–471 (2009).

40. Speich, S., Blanke, B. & Cai, W. Atlantic meridional overturning circulation and the Southern Hemisphere supergyre. *Geophys. Res. Lett.* **34,** L23614 (2007).

41. Gordon, A. L. Interocean exchange of thermocline water. *J. Geophys. Res.* **91,** 5037–5046 (1986).

42. Rintoul, S. R. South Atlantic interbasin exchange. *J. Geophys. Res.* **96,** 2675–2692 (1991).

43. Sloyan, B. M. & Rintoul, S. R. The Southern Ocean limb of the global deep overturning circulation. *J. Phys. Oceanogr.* **31,** 143–173 (2001).

44. Van Sebille, E. & van Leeuwen, P. J. Fast northward energy transfer in the Atlantic due to Agulhas Rings. *J. Phys. Oceanogr.* **37,** 2305–2315 (2007).

45. Cunningham, S. A. *et al.* Temporal variability of the Atlantic meridional overturning circulation at 26.5° N. *Science* **317,** 935–938 (2007).

46. Schouten, M. W., De Ruijter, W. P. M. & Van Leeuwen, P. J. Upstream control of Agulhas ring shedding. *J. Geophys. Res.* **107,** 3109 (2002).
    This work is an observational study (satellite altimetry data) showing remote control of Agulhas ring formation by eddies from the Mozambique Channel and EMC, which in turn are triggered by westwards-propagating Rossby waves along 10° S and 24° S. Interannual modulations seem to be connected to the Indian Ocean Dipole.

47. Haarsma, R. J., Campos, E. J. D., Drijfhout, S., Hazeleger, W. & Severijns, C. Impacts of interruption of the Agulhas Leakage on the tropical Atlantic in coupled ocean-atmosphere simulations. *Clim. Dyn.* **36,** 989–1003 (2009).

48. Hughes, G. O., Hogg, A., McC. & Griffiths, R. W. Available potential energy and irreversible mixing in the meridional overturning circulation. *J. Phys. Oceanogr.* **39,** 3130–3146 (2009).

49. Toggweiler, J. R. & Samuels, B. Effect of Drake Passage on the global thermohaline circulation. *Deep-Sea Res. I* **42,** 477–500 (1995).

50. Sijp W. P. & England, M. H. Southern hemisphere westerly wind control over the ocean's thermohaline circulation. *J. Clim.* **22,** 1277–1286 (2009).

51. Sen Gupta, A. *et al.* Projected changes to the Southern Hemisphere ocean and sea-ice in the IPCC AR4 climate models. *J. Clim.* **22,** 3047–3078 (2009).

52. Toggweiler, J. R., Russell, J. L. & Carson, S. R. Midlatitude westerlies, atmospheric CO_2, and climate change during the ice ages. *Paleoceanography* **21,** PA2005 (2006).

53. Watson, A. J. & Naviera Garabato, A. C. The role of Southern Ocean mixing and upwelling in glacial-interglacial atmospheric CO_2 change. *Tellus* **58B,** 73–87 (2006).

54. Tschumi, T., Joos, F. & Parekh, P. How important are Southern Hemisphere wind changes for low glacial carbon dioxide? A model study. *Paleoceanography* **23,** PA4208 (2008).

55. Speer, K., Rintoul, S. R. & Sloyan, B. The diabatic deacon cell. *J. Phys. Oceanogr.* **30,** 3212–3222 (2000).

56. Naviera Garabato, A. C., Stevens, D. P., Watson, A. J. & Roether, W. Short-circuiting of the overturning circulation in the Antarctic Circumpolar Current. *Nature* **447,** 194–197 (2007).

57. Farneti, R. & Delworth, T. L. The role of mesoscale eddies in the remote oceanic response to altered Southern Hemisphere winds. *J. Phys. Oceanogr.* **40,** 2348–2354 (2010).

58. Jury, M. & Walker, N. Marine boundary layer modification across the edge of the Agulhas Current. *J. Geophys. Res.* **93,** 647–654 (1988).

59. Rouault, M., Lee-Thorp, A. M. & Lutjeharms, J. R. E. The atmospheric boundary layer above the Agulhas Current during along-current winds. *J. Phys. Oceanogr.* **30,** 40–50 (2000).

60. Liu, W. T., Xie, X. & Niiler, P. P. Ocean–atmosphere interaction over Agulhas Extension meanders. *J. Clim.* **20,** 5784–5797 (2007).

61. Reason, C. J. C. Evidence for the influence of the Agulhas Current on REGIONAL atmospheric circulation patterns. *J. Clim.* **14,** 2769–2778 (2001).

62. Nakamura, H. & Shimpo, A. Seasonal variations in the Southern Hemisphere storm tracks and jet streams as revealed in reanalysis datasets. *J. Clim.* **17,** 1828–1844 (2004).

63. Rouault, M., White, S. A., Reason, C. J. C., Lutjeharms, J. R. E. & Jobard, I. Ocean-atmosphere interaction in the Agulhas Current region and a South African extreme weather event. *Weather Forecast.* **17,** 655–669 (2002).

64. Behera, S. K. & Yamagata, T. Subtropical SST dipole events in the southern Indian Ocean. *Geophys. Res. Lett.* **28,** 327–330 (2001).

65. Hermes, J. C. & Reason, C. J. C. Ocean model diagnosis of interannual coevolving SST variability in the South Indian and South Atlantic oceans. *J. Clim.* **18,** 2864–2882 (2005).

66. Zinke, J., Dullo, W.-C., Heiss, G. A. & Eisenhauer, A. ENSO and Indian Ocean subtropical dipole variability is recorded in a coral record off southwest Madagascar for the period 1659 to 1995. *Earth Planet. Sci. Lett.* **228,** 177–194 (2004).

67. Palastanga, V., van Leeuwen, P. J. & De Ruijter, W. P. M. A link between low-frequency mesoscale eddy variability around Madagascar and the large-scale Indian Ocean variability. *J. Geophys. Res.* **111**, C09029 (2006).

68. Beal, L. M. A time series of Agulhas Undercurrent transport. *J. Phys. Oceanogr.* **39**, 2436–2450 (2009).

69. Biastoch, A., Beal, L. M., Casal, T. G. D. & Lutjeharms, J. R. E. Variability and coherence of the Agulhas Undercurrent in a high-resolution ocean general circulation model. *J. Phys. Oceanogr.* **39**, 2417–2435 (2009).

70. Tsugawa, M. & Hasumi, H. Generation and growth mechanism of a Natal Pulse. *J. Phys. Oceanogr.* **40**, 1597–1612 (2010).

71. Rau, A. J., Rogers, J. & Chen, M.-T. Late Quaternary palaeoceanographic record in giant piston cores off South Africa, possibly including evidence of neotectonism. *Quaternary Int.* **148**, 65–77 (2006).

72. Lisiecki, L. E., Raymo, M. E. & Curry, W. B. Atlantic overturning responses to Late Pleistocene climate forcings. *Nature* **456**, 85–88 (2008).

73. Anderson, R. F. *et al.* Wind-driven upwelling in the Southern Ocean and the deglacial rise in atmospheric CO_2. *Science* **323**, 1443–1448 (2009).

74. Thornalley, D. J. R., Elderfield, H. & McCave, N. Reconstructing North Atlantic deglacial surface hydrography and its link to the Atlantic overturning circulation. *Global Planet. Change.* (in the press).

75. Zahn, R. Beyond the CO_2 connection. *Nature* **460**, 335–336 (2009).

76. Cai, W. Antarctic ozone depletion causes an intensification of the Southern Ocean super-gyre circulation. *Geophys. Res. Lett.* **33**, L03712 (2006).

77. Vellinga, M. & Wood, R. A. Global climatic impacts of a collapse of the Atlantic thermohaline circulation. *Clim. Change* **54**, 251–267 (2005).

78. Stephens, B. B. & Keeling, R. F. The influence of Antarctic sea ice on glacial-interglacial CO_2 variations. *Nature* **404**, 171–174 (2000).

79. Wunsch, C. & Ferrari, R. Vertical mixing, energy, and the general circulation of the oceans. *Annu. Rev. Fluid Mech.* **36**, 281–314 (2004).

80. Kuhlbrodt, T. *et al.* On the driving processes of the Atlantic meridional overturning circulation. *Rev. Geophys.* **45**, RG2001 (2007).

81. Weijer, W., De Ruijter, W. P. M., Dijkstra, H. A. & van Leeuwen, P. J. Impact of interbasin exchange on the Atlantic overturning circulation. *J. Phys. Oceanogr.* **29**, 2266–2284 (1999).

82. McDonagh, E. L. & King, B. A. Oceanic fluxes in the South Atlantic. *J. Phys. Oceanogr.* **35**, 109–122 (2005).

83. Rahmstorf, S. On the freshwater forcing and transport of the Atlantic thermohaline circulation. *Clim. Dyn.* **12**, 799–811 (1996).

84. DeVries, P. & Weber, S. L. The Atlantic freshwater budget as a diagnostic for the existence of a stable shut down of the meridional overturning circulation. *Geophys. Res. Lett.* **32**, L09606 (2005).

85. Dijkstra, H. A. Characterization of the multiple equilibria regime in a global ocean model. *Tellus* **59A**, 695–705 (2007).

86. Romanova, V., Kohl, A., Stammer, D., Klepp, C. & Andersson, A. Sea surface freshwater flux estimates from GECCO, HOAPS and NCEP. *Tellus* **62**, 435–452 (2010).

87. Huisman, S. E., den Toom, M., Dijkstra, H. A. & Drijfhout, S. An indicator of the multiple equilibria regime of the Atlantic meridional overturning circulation. *J. Phys. Oceanogr.* **40**, 551–567 (2010).

88. Marsh, R., Hazeleger, W., Yool, A. & Rohling, E. J. Stability of the thermohaline circulation under millennial CO_2 forcing and two alternative controls on Atlantic salinity. *Geophys. Res. Lett.* **34**, L03605 (2007).

89. Cortese, G. & Abelmann, A. Radiolarian-based paleotemperatures during the last 160 kyr at ODP Site 1089 (Southern Ocean, Atlantic Sector). *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **182**, 259–286 (2002).

90. Martinez-Mendez, G. *et al.* Contrasting multi-proxy reconstructions of surface ocean hydrography in the Agulhas Corridor and implications for the Agulhas Leakage during the last 345,000 years. *Paleoceanography* **25**, PA4227 (2010).

91. Franzese, A. M., Hemming, S. R. & Goldstein, S. L. Use of strontium isotopes in detrital sediments to constrain the glacial position of the Agulhas Retroflection. *Paleoceanography* **24**, PA2217 (2009).

92. Large, W. G. & Yeager, S. G. *Diurnal to Decadal Global Forcing for Ocean and Sea-Ice Models: The Data Sets and Flux Climatologies.* NCAR Tech. Note NCAR/TN-460+STR (NCAR, 2004).

93. Donlon, C. *et al.* The global ocean data assimilation experiment high-resolution sea surface temperature pilot project. *Bull. Am Meteorol. Soc.* **88**, 1197–1213 (2007).

94. Mix, A. C., Le, J. & Shackleton, N. J. Benthic foraminiferal stable isotope stratigraphy of Site 846: 0–1.8 Ma. *Proc. ODP Sci. Results* **138**, 839–854 (1995).

95. Speich, S., Lutjeharms, J. R. E., Penven, P. & Blanke, B. Role of bathymetry in Agulhas Current configuration and behaviour. *Geophys. Res. Lett.* **33**, L23611 (2006).

96. Zharkov, V. & Nof, D. Retroflection from slanted coastlines – circumventing the ''vorticity paradox''. *Ocean Sci. Discuss.* **4**, 293–306 (2008).

97. Dijkstra, H. A. & de Ruijter, W. P. M. On the physics of the Agulhas Current: steady retroflection regimes. *J. Phys. Oceanogr.* **31**, 2971–2985 (2001).

98. Van Sebille, E., Biastoch, A., van Leeuwen, P. J. & De Ruijter, W. P. M. A weaker Agulhas Current leads to more Agulhas Leakage. *Geophys. Res. Lett.* **36**, L03601 (2009).

99. Rojas, M. *et al.* The Southern Westerlies during the last glacial maximum in PMIP2 simulations. *Clim. Dyn.* **32**, 525–548 (2009).

100. Banks, H. T., Stark, S. & Keen, A. B. The adjustment of the coupled climate model HadGEM1 toward equilibrium and the impact on global climate. *J. Clim.* **20**, 5815–5826 (2007).

Author Contributions The ideas presented here were developed jointly by L.M.B., W.P.M.D.R., A.B., R.Z. and the SCOR/WCRP/IAPSO Working Group 136 on the Climatic Importance of the Greater Agulhas System. Working Group 136 is sponsored by the Scientific Committee for Oceanic Research (SCOR), the World Climate Research Program (WCRP) and the International Association for the Physical Sciences of the Ocean (IAPSO).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to L.M.B. (lbeal@rsmas.miami.edu).

SCOR/WCRP/IAPSO Working Group 136

Meghan Cronin[1], Juliet Hermes[2], Johann Lutjeharms[3], Graham Quartly[4], Tomoki Tozuka[5], Sheekela Baker-Yeboah[6], Thomas Bornman[7], Paolo Cipollini[4], Henk Dijkstra[8], Ian Hall[9], Wonsun Park[10], Frank Peeters[11], Pierrick Penven[12], Herman Ridderinkhof[13] & Jens Zinke[13]

[1]NOAA Pacific Marine Environmental Laboratory, 7600 Sand Point Way NE, Seattle, Washington 98115, USA. [2]South African Environmental Observation Network, Private Bag X2, Roggebaai 8012, South Africa. [3]Department of Oceanography, University of Cape Town, 7700 Rondebosch, South Africa. [4]National Oceanography Centre, Southampton, European Way, Southampton SO14 3ZH, UK. [5]Department of Earth and Planetary Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. [6]Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA. [7]South African Institute for Aquatic Biodiversity, Private Bag 1015, Grahamstown 6140, South Africa. [8]Institute for Marine and Atmospheric Research, Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands. [9]School of Earth and Ocean Sciences, Cardiff University, Park Place, Cardiff CF10 3AT, UK. [10]Leibniz-Institut für Meereswissenschaften (IFM-GEOMAR), Düsternbrooker Weg 20, 24105 Kiel, Germany. [11]Faculty of Earth and Life Sciences, Section Marine Biogeology, Vrije Universiteit Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands. [12]Laboratoire de Physique des Oceans (UMR 6523 CNRS, IFREMER, IRD, UBO), LMI ICEMASA, Department of Oceanography, University of Cape Town, Cape Town, 7700 Rondebosch, South Africa. [13]NIOZ Royal Netherlands Institute for Sea Research, PO Box 59, 1790 AB Den Burg, The Netherlands.

ARTICLE

Shank3 mutant mice display autistic–like behaviours and striatal dysfunction

João Peça[1,2]*, Cátia Feliciano[1,3]*, Jonathan T. Ting[1], Wenting Wang[1], Michael F. Wells[1], Talaignair N. Venkatraman[4], Christopher D. Lascola[1,4], Zhanyan Fu[1,5,6] & Guoping Feng[1,6,7]

Autism spectrum disorders (ASDs) comprise a range of disorders that share a core of neurobehavioural deficits characterized by widespread abnormalities in social interactions, deficits in communication as well as restricted interests and repetitive behaviours. The neurological basis and circuitry mechanisms underlying these abnormal behaviours are poorly understood. SHANK3 is a postsynaptic protein, whose disruption at the genetic level is thought to be responsible for the development of 22q13 deletion syndrome (Phelan–McDermid syndrome) and other non-syndromic ASDs. Here we show that mice with *Shank3* gene deletions exhibit self-injurious repetitive grooming and deficits in social interaction. Cellular, electrophysiological and biochemical analyses uncovered defects at striatal synapses and cortico-striatal circuits in *Shank3* mutant mice. Our findings demonstrate a critical role for SHANK3 in the normal development of neuronal connectivity and establish causality between a disruption in the *Shank3* gene and the genesis of autistic-like behaviours in mice.

Autism and autism spectrum disorders (ASDs) are neurodevelopmental disorders diagnosed based on a triad of criteria: deficits in communication, impaired social interaction, and repetitive or restricted interests and behaviours[1]. ASDs are highly heritable disorders with concordance rates as high as 90% for monozygotic twins[2]. Recent genetic and genomic studies have identified a large number of candidate genes for ASDs[3], many of which encode synaptic proteins[4–6], indicating synaptic dysfunction may have a critical role in ASDs[7,8]. One of the most promising ASD candidate genes is *Shank3*, which codes for a key postsynaptic density (PSD) protein at glutamatergic synapses. Disruption of *Shank3* is thought to be the cause of core neurodevelopmental and neurobehavioural deficits in the 22q13 deletion syndrome (Phelan–McDermid syndrome), an autism spectrum disorder[9–11]. Furthermore, recent genetic screens have identified several mutations/rare variants of the *Shank3* gene in ASD patients outside of diagnosed 22q13 deletion syndrome[12,13].

The Shank family of proteins (SHANK1–3) directly bind SAPAP (also known as DLGAP) to form the PSD-95–SAPAP–SHANK complex[14,15] (PSD-95 is also known as DLG4). This core of proteins is thought to function as a scaffold, orchestrating the assembly of the macromolecular postsynaptic signalling complex at glutamatergic synapses. Currently, however, little is known about the *in vivo* function of SHANK3 at the synapse and how a disruption of *Shank3* may contribute to ASDs. Here we demonstrate that genetic disruption of *Shank3* in mice leads to compulsive/repetitive behaviour and impaired social interaction, resembling two of the cardinal features of ASDs. Biochemical, morphological and electrophysiological studies revealed synaptic dysfunction at cortico-striatal synapses, part of the neural circuits strongly implicated as dysfunctional in ASDs. Our studies provide a synaptic and circuitry mechanism underlying *Shank3* disruption and ASD-like behaviours.

Shank3B⁻/⁻ mice display repetitive grooming

The *Shank3* gene codes for large proteins with multiple protein–protein interaction domains (Fig. 1a). We generated two different alleles of SHANK3 mutant mice. In *Shank3A* mutant mice, we targeted a portion of the gene encoding the ankyrin repeats (Supplementary Fig. 1b). This resulted in a complete elimination of SHANK3$_\alpha$, the longest SHANK3 isoform (Fig. 1b). However, the other two isoforms were not affected (here named SHANK3$_\beta$ and SHANK3$_\gamma$). In *Shank3B* mutants, we targeted the fragment encoding the PDZ domain (Supplementary Fig. 1c). This led to the complete elimination of both SHANK3$_\alpha$ and SHANK3$_\beta$ isoforms and a significant reduction of the putative SHANK3$_\gamma$ isoform at the PSD ($-42.12\% \pm 9.27\%$ of control, $n = 3$, $P < 0.05$) (Fig. 1b). Our analysis is mainly focused on the *Shank3B*⁻/⁻ mutants due to their more pronounced behavioural and physiological defects.

We used mice with a hybrid genetic background to avoid the potential contribution to behavioural phenotypes of homozygous genetic variants on a pure inbred background[7,16]. Initially, F1 hybrids from heterozygous × heterozygous matings were generated and homozygous mice were born at an expected Mendelian rate. However, homozygous knockout mice from this type of mating are smaller than their wild-type littermates, presumably due to inadequate competition for resources during early postnatal days leading to different developmental trajectories. We postulated that this size difference would influence our behavioural tests. To alleviate this confound, heterozygous animals were crossed in direct brother-sister matings for five generations from which we derived F5 isogenic hybrids in a mixed background. These isogenic animals were then used to generate time-mated homozygous × homozygous breeding pairs to obtain wild-type and mutant animals used in the experiments. F5 *Shank3A* and F5 *Shank3B* knockouts from these matings are reared to weaning age with body weights similar to those from F5 control animals.

Shank3B⁻/⁻ mice did not display any gross anatomical or histological brain abnormality, but on rare occasions exhibited seizures during handling in routine husbandry procedures. However, spontaneous seizures were never observed. By the age of 3–6 months, *Shank3B*⁻/⁻

[1]Department of Neurobiology, Duke University Medical Center, Durham, North Carolina 27710, USA. [2]PhD Programme in Biomedicine and Experimental Biology (BEB), Center for Neuroscience and Cell Biology, University of Coimbra, Coimbra, Portugal. [3]Gulbenkian PhD Programme in Biomedicine, Gulbenkian Science Institute, 2781-901 Oeiras, Portugal. [4]Department of Radiology, and Brain Imaging and Analysis Center, Duke University Medical Center, Durham, North Carolina 27710, USA. [5]Department of Psychiatry and Behavioral Science, Duke University Medical Center, Durham, North Carolina 27710, USA. [6]McGovern Institute for Brain Research, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [7]Stanley Center for Psychiatric Research, Broad Institute, Cambridge, Massachusetts 02142, USA.
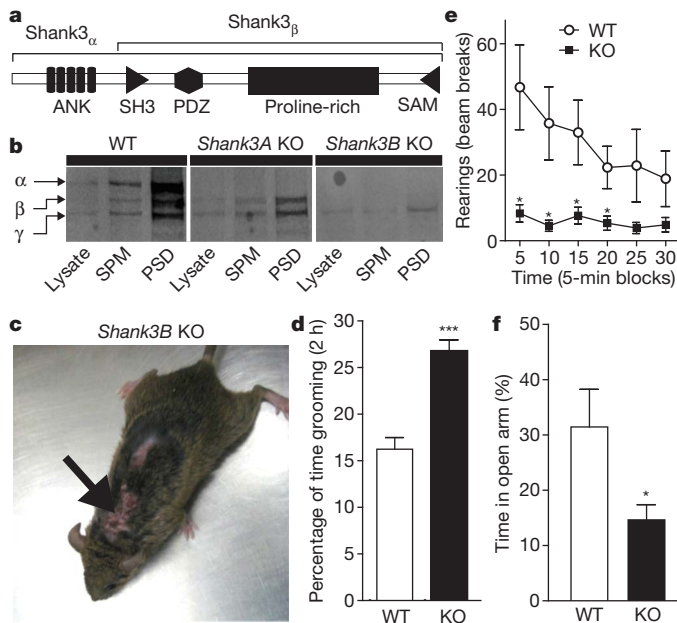*These authors contributed equally to this work.

Figure 1 | Excessive grooming, skin lesions and anxiety-like behaviour in _Shank3B_$^{-/-}$ mice. a, SHANK3 protein structure. **b,** Western blot showing a pan-SHANK3 antibody staining in brain lysate, synaptosomal plasma membrane (SPM) and 2× Triton X-100-washed PSD (PSD) fraction in wild-type (WT), _Shank3A_$^{-/-}$ and _Shank3B_$^{-/-}$ mice. **c,** Four-month-old _Shank3B_$^{-/-}$ mice display neck and head lesions (arrows). **d,** Pre-lesion _Shank3B_$^{-/-}$ (KO) mice spent more time in self-grooming than WT. **e,** In the open field test, _Shank3B_$^{-/-}$ mice, when compared to controls, display decreased rearing activity. **f,** In the zero maze test, _Shank3B_$^{-/-}$ mice spent less time in the open area than wild-type controls. *$P < 0.05$, ***$P < 0.001$, two-tailed t-test for **d** and **f**, two-way repeated measures ANOVA with post hoc two-tailed t-test for **e**; all data are presented as means ± s.e.m. from 6–9 mice per genotype.

mice developed pronounced skin lesions with varying degrees of phenotypical penetrance: approximately 35% in the general holding colony (Fishers exact test, $P < 0.0001$), and 100% in mating females that have produced 4–6 litters. The lesions tend to appear first on the back of the neck or on the face (Fig. 1c) and usually progressed bilaterally to cover large areas of the body. The lesions were self-inflicted, as they were present in animals socially isolated at weaning age, and not due to excessive allogrooming, as no lesions were found in wild-type or _Shank3B_$^{+/-}$ mice housed from birth with _Shank3B_$^{-/-}$ animals. Furthermore, 24 h videotaping in pre-lesion animals revealed that _Shank3B_$^{-/-}$ mice showed an increase in time spent grooming when compared to wild-type controls (Fig. 1d). These observations indicate that _Shank3B_$^{-/-}$ mice display excessive grooming and self-injurious behaviour.

We characterized the animals further in a battery of behavioural tests. In the rotarod motor test, _Shank3B_$^{-/-}$ and control animals performed at similar levels (Supplementary Fig. 2). In the open field test, when compared to controls, _Shank3B_$^{-/-}$ mice showed similar levels of activity and thigmotaxis (Supplementary Fig. 2). However, rearing, which is a form of vertical exploration considered to be anxiogenic for mice, was significantly reduced in the mutants (Fig. 1e). In the elevated zero maze, the _Shank3B_$^{-/-}$ mice spent less time exploring the open arms of the maze versus the closed arms (Fig. 1f). In the light-dark emergence test, the _Shank3B_$^{-/-}$ mice displayed an increased latency to cross into the brightly lit area, although the time spent in each side of the box was similar between mutant animals and controls (Supplementary Fig. 2). Thus, the _Shank3B_$^{-/-}$ mice display an anxiety-like behaviour and excessive, self-injurious grooming. In contrast, _Shank3A_$^{-/-}$ mice displayed no lesions or anxiety-like behaviour (Supplementary Fig. 3).

Social interaction deficits in _Shank3B_$^{-/-}$ mice

Deficits in social interaction are the most recognizable manifestation of autistic behaviours in humans. We used a modified version of a three-chamber social arena[17] to probe animals for their voluntary initiation of social interaction and their ability to discriminate social novelty. Initially, the test animal was left to explore and initiate social contact with a partner ('Stranger 1') held inside a wired cage or an identical but empty wired cage ('Empty cage'). In this test, the _Shank3B_$^{-/-}$ mice displayed dysfunctional social interaction behaviour, as measured by observing both time spent in the compartment containing the social partner (Fig. 2a, b) or in close interaction (Fig. 2d). Notably, _Shank3B_$^{-/-}$ mice exhibited a clear preference for interacting with the empty cage rather than with the social partner (Fig. 2a, d). In a subsequent trial, a novel social partner ('Stranger 2') was introduced into the previously empty wired cage. Wild-type mice displayed a preference for the novel animal, as shown by the increase in time spent in the compartment containing 'Stranger 2'. The _Shank3B_$^{-/-}$ mutants markedly spent more time in the centre chamber (Fig. 2c) and a reduced amount of time closely interacting with either social partner (Fig. 2e). In an identical test, the _Shank3A_$^{-/-}$ mice displayed normal initiation of social interaction, but perturbed recognition of social novelty (Supplementary Fig. 4).

Additionally, in an open arena test, freely interacting dyadic pairs of wild-type–_Shank3B_$^{-/-}$ mice displayed less time spent in reciprocal interaction, a lower frequency of nose-to-nose interaction and anogenital sniffing when compared to wild-type–wild-type pairs (Supplementary Fig. 5). Thus, data from both social interaction tests indicate that _Shank3B_$^{-/-}$ mice display abnormal social interaction as well as deficits in discriminating social novelty.

In our breeding scheme, _Shank3B_$^{-/-}$ mice and wild-type mice were nurtured by _Shank3B_$^{-/-}$ and wild-type dams, respectively. To assess the impact of maternal rearing on the observed sociability defects, we performed time-mated cross-fostering of _Shank3B_$^{-/-}$ mice and controls. Cross-fostering of _Shank3B_$^{-/-}$ neonatal pups with wild-type dams (KO$_{cf}$) revealed qualitatively equivalent social defects in the mutant mice as compared to those observed in mutant mice nurtured by _Shank3B_$^{-/-}$ dams. Additionally, rearing wild-type neonatal pups by _Shank3B_$^{-/-}$ dams (WT$_{cf}$) did not perturb normal sociability in wild-type animals (Supplementary Fig. 6). These data further indicate a genetic origin of the abnormal social behaviours in the _Shank3_ mutant mice.

Altered PSD composition in the striatum

The basal ganglia are one of the brain regions implicated as dysfunctional in ASD. The repetitive/compulsive grooming behaviour in _Shank3B_$^{-/-}$ mice also suggests defects in cortico-striatal function. Furthermore, _Shank3_, but not _Shank1_ or _Shank2_, is highly expressed in the striatum (Fig. 3a) (Supplementary Fig. 7 and Supplementary Table 1). Therefore, we focused our analyses on striatal neurons and cortico-striatal synapses.

Shank family members have been proposed as key regulators of the PSD at glutamatergic synapses[18]. To determine how the disruption of _Shank3_ may affect the PSD protein network, we used biochemically purified PSDs from the striatum of wild-type and _Shank3B_$^{-/-}$ mice and performed semiquantitative western blotting for several scaffolding proteins (Fig. 3b) and glutamate receptor subunits (Fig. 3c). At the PSD level, we observed reduced levels of SAPAP3, Homer-1b/c and PSD-93 (also known as HOMER1 and DLG2, respectively; Fig. 3b) as well as a reduction in the glutamate receptor subunits GluR2, NR2A and NR2B (also known as GRIA2, GRIN2A and GRIN2B, respectively; Fig. 3c). These results suggest an altered molecular composition of postsynaptic machinery in the striatum and a possible disruption of glutamatergic signalling.

Morphological defects of medium spiny neurons

To test whether disruption of SHANK3 affects neuronal morphology, we traced Golgi-stained striatal medium spiny neurons (MSNs) and
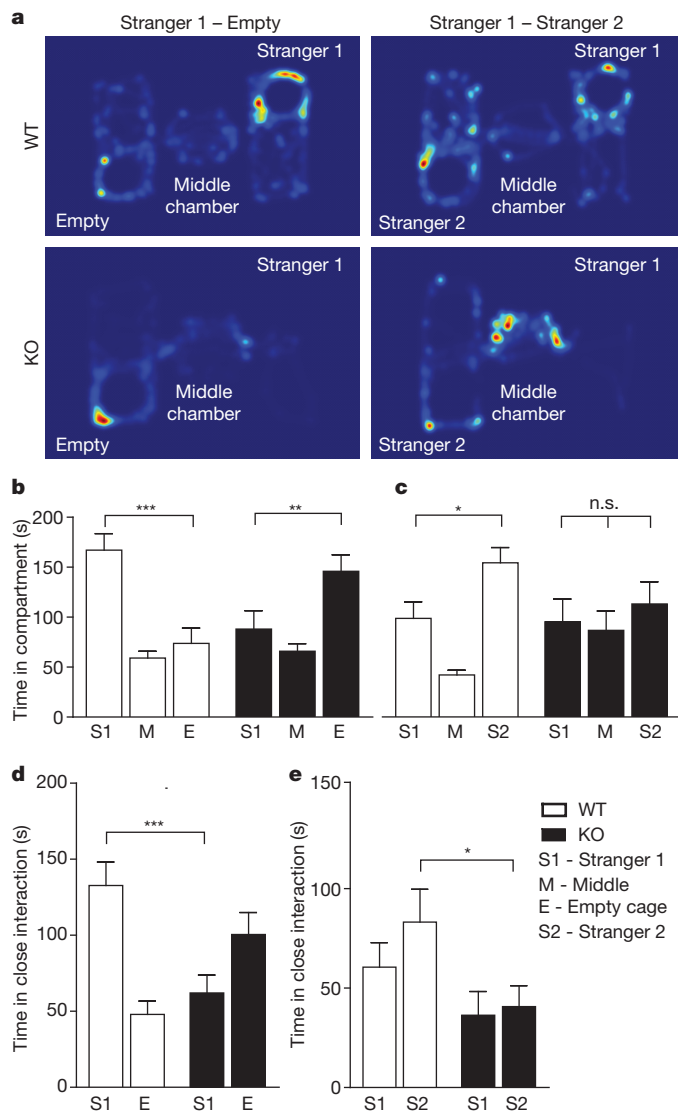
Figure 2 | Reduced social interaction and abnormal social novelty recognition in Shank3B$^{-/-}$ mice. a, Representative heat map analysis from 'Stranger 1 – Empty' and 'Stranger 1 – Stranger 2' trials from Shank3B$^{-/-}$ mice and controls. **b**, In the social interaction test, Shank3B$^{-/-}$ mice (closed bars) spent less time in the chamber containing the social partner (Stranger 1) and more time in the chamber containing the empty wire cage when compared to controls (open bars). **c**, In the social novelty test, Shank3B$^{-/-}$ mice do not display a preference for the novel social partner (Stranger 2), and spent more time in the middle chamber. **d**, **e**, When analysing social interaction by close proximity (within 5 cm) to either 'Stranger 1', 'Empty Cage' (**d**), or 'Stranger 1', 'Stranger 2' (**e**), Shank3B$^{-/-}$ mice displayed a clear reduction in social interaction when compared to controls (**d**); whereas under a social novelty paradigm (**e**), Shank3B$^{-/-}$ mice displayed a clear reduction in time spent with 'Stranger 2' *$P<0.05$, **$P<0.01$, ***$P<0.0001$; one-way ANOVA, with Bonferroni post hoc t-test for **b**–**e**; all data presented as means ± s.e.m.; 12–14 mice per group.

their dendrites to investigate the cellular morphology and complexity of these cells. Sholl analysis revealed neuronal hypertrophy as measured by an increase in complexity of dendritic arborizations (Fig. 4a), total dendritic length (Fig. 4b) and also an increase in surface area (Fig. 4c) in Shank3B$^{-/-}$ MSNs.

Next, we performed patch-assisted Lucifer Yellow cell filling of MSNs and measured spine density in control and Shank3B$^{-/-}$ mice. Shank3B$^{-/-}$ mice displayed a significant reduction in spine density (Fig. 4d, e). We did not observe significant changes in spine length or head diameter; however, the neck width of Shank3B$^{-/-}$ MSN spines was slightly larger than that of controls (Supplementary Fig. 8).
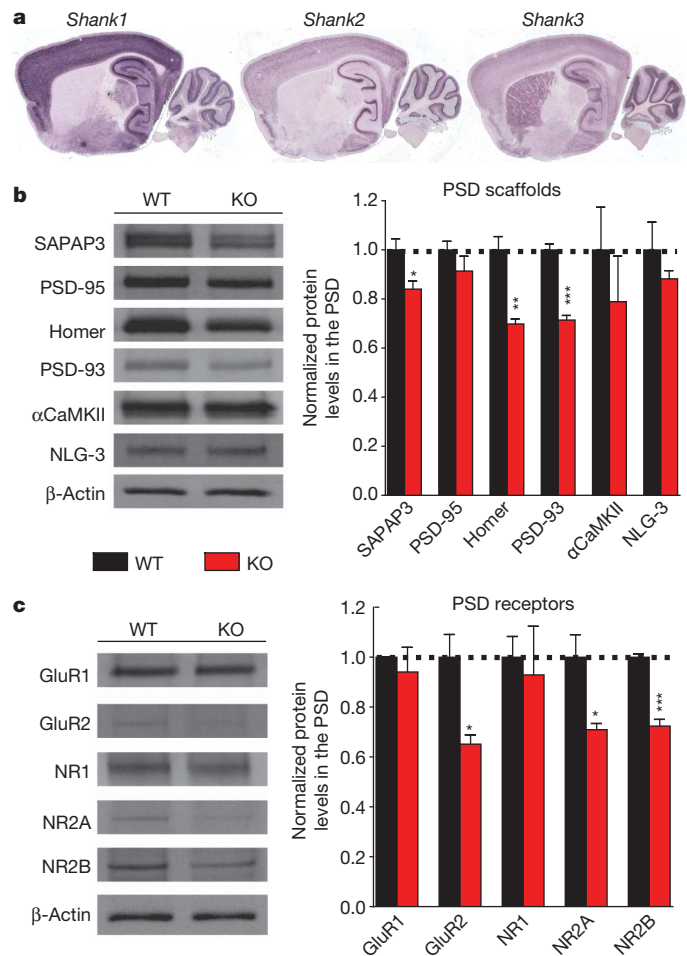

Figure 3 | Biochemical changes in striatal synapses of Shank3B$^{-/-}$ mice. a, Only Shank3 mRNA is highly expressed in the striatum. **b**, Protein levels of the scaffolding proteins SAPAP3, Homer and PSD-93 are reduced in striatal PSD fractions from Shank3B$^{-/-}$ mice. αCaMKII and NLG-3 are also known as CAMK2A and NLGN3, respectively. **c**, Protein levels of glutamate receptor subunits GluR2, NR2A and NR2B are reduced in striatal PSD fractions from Shank3B$^{-/-}$ mice. GluR1 and NR1 are also known as GRIA1 and GRIN1, respectively. Each lane was loaded with 3 μg of protein with β-actin as loading control and normalized to wild-type levels. *$P<0.05$, **$P<0.01$, ***$P<0.001$, two-tailed t-test; all data are presented as means ± s.e.m.; $n=3$ samples per group, with each sample being a combined pool of striatal tissue from three animals.

Finally we analysed PSD morphology by electron microscopy (Fig. 4f). We found a significant reduction in mean thickness (Fig. 4g) of PSDs from Shank3B$^{-/-}$ mice relative to controls. Additionally, PSD length was also significantly reduced in the Shank3B$^{-/-}$ mice (Fig. 4h). Taken together, these results highlight a critical in vivo role for Shank3 in the normal development of medium spiny neurons and striatal glutamatergic synapses.

Striatal hypertrophy in Shank3B$^{-/-}$ mice

Even though there is no clear correlation between brain size or neuronal hypertrophy specifically for Shank3 disruptions in humans, a potential link between enlarged brain size, neuronal hypertrophy and autism has been suggested previously[19]. In particular, increased caudate volume in autism patients has been proposed to be linked to repetitive behaviours[20,21]. We measured striatal volume using three-dimensional magnetic resonance imaging in the intact brain of Shank3B$^{-/-}$ and control mice. We found that there was no significant difference in overall brain size between the genotypes. However, measurement of caudate volume in the same animals revealed a small but
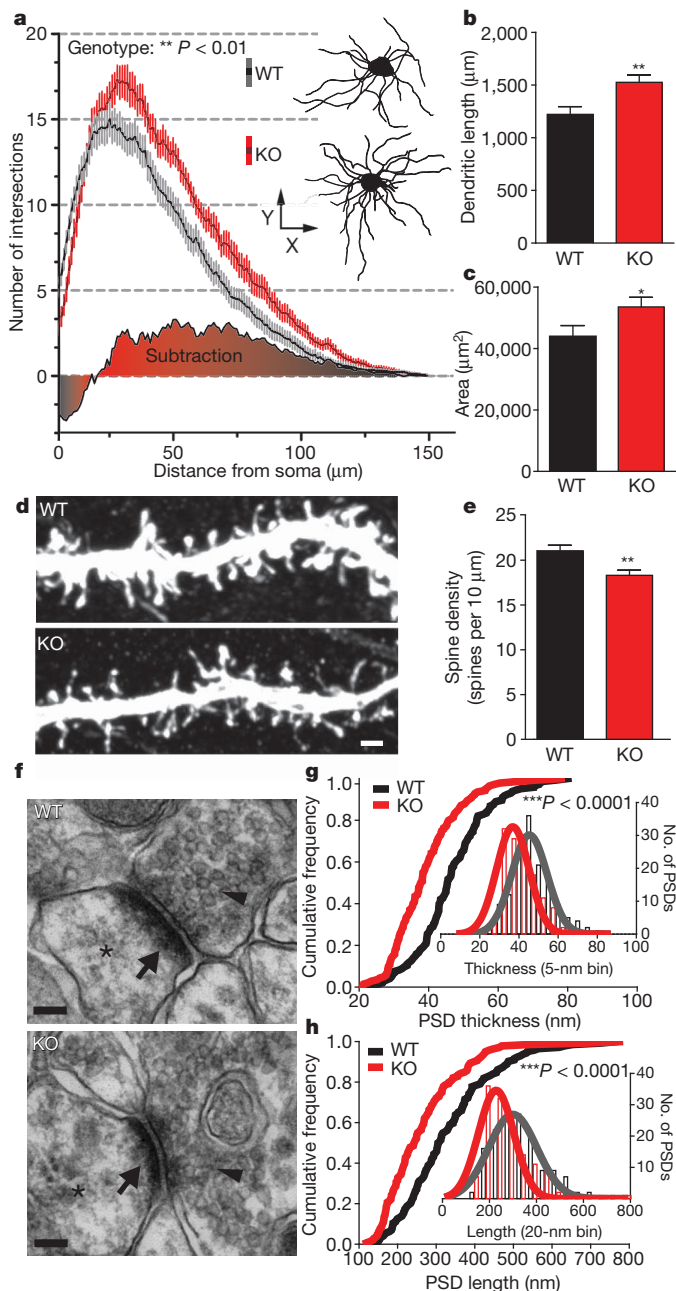
significant volumetric enlargement of this structure in $Shank3B^{-/-}$ mice (Supplementary Fig. 9). These data suggest a correlation between neuronal hypertrophy and brain volume, consistent with studies from other mouse models of ASD[22,23].

Perturbation of striatal postsynaptic function

To elucidate the functional consequences of a disruption in $Shank3$ on synaptic function, we performed recordings of cortico-striatal synaptic circuitry in acute brain slices of 6–7-week-old animals. We found that field population spikes were significantly reduced in $Shank3B^{-/-}$ mice when compared with controls (Fig. 5a). Presynaptic function was not altered, as indicated by the relationship of stimulation intensity to the amplitude of the action potential component of the response termed negative peak 1 (NP1) and the paired-pulse ratio (PPR; Supplementary Fig. 10). These results indicate that the reduction in total field responses was most likely due to a postsynaptic impairment in synaptic function and/or a reduction in the number of functional synapses. Consistent with their mild behavioural phenotypes $Shank3A^{-/-}$ mice displayed minimal disruption at cortico-striatal synapses (Supplementary Fig. 11).
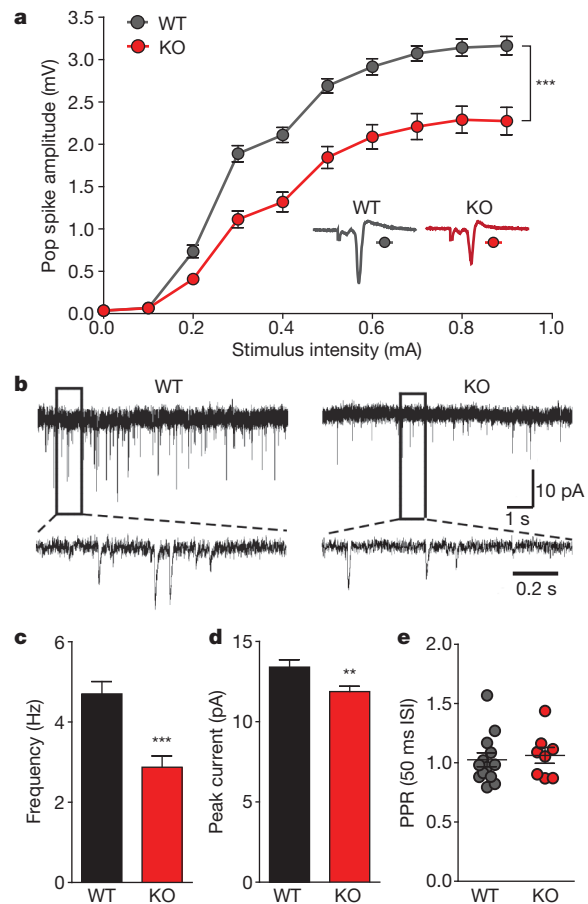

Figure 4 | Morphological and ultrastructural neuronal abnormalities in $Shank3B^{-/-}$ mice. a, Sholl analysis reveals an increased neuronal complexity of $Shank3B^{-/-}$ MSNs (red) when compared to MSNs from wild-type mice (grey); example neurons are shown as insets (top, WT–; bottom, KO–). **b, c**, MSNs from $Shank3B^{-/-}$ mice show an increase in total dendritic length (**b**) and surface area (**c**) when compared to controls. **d**, Representative confocal stacks of dye-filled MSNs from KO and WT mice; scale bar; 1 μm. **e**, Spine density in MSNs from $Shank3B^{-/-}$ mice is lower than that of wild-type MSNs. **f**, Examples of electron micrographs depicting the synaptic contacts with presynaptic vesicles (arrowheads), postsynaptic densities (arrow) and dendritic spine (asterisk); scale bar, 100 nm. **g**, $Shank3B^{-/-}$ PSDs are thinner than wild-type PSDs. **h**, $Shank3B^{-/-}$ PSDs are shorter than wild-type PSDs. *$P < 0.05$, **$P < 0.01$, ***$P < 0.0001$; two-way repeated measures ANOVA for **a**; two-tailed t-test for **b**, **c** and **e**; two-sample Kolmogorov–Smirnov test for **g** and **h**. Data in **g** and **h** are presented as cumulative frequency plot with histogram distribution and Gaussian curve fit for the insets. Data from **b**, **c** and **e** are presented as means ± s.e.m.; $n = 36$ from 3 wild-type mice and $n = 36$ from 3 $Shank3B^{-/-}$ mice for **a–c**; $n = 41$ dendritic segments from 3 wild-type mice and $n = 36$ dendritic segments from 3 $Shank3B^{-/-}$ mice for **e**; $n = 144$ PSDs from three wild-type mice and $n = 140$ PSDs from three $Shank3B^{-/-}$ mice for **g**, **h**.

Figure 5 | Reduced cortico-striatal synaptic transmission in $Shank3B^{-/-}$ MSNs. a, Cortico-striatal pop spike amplitude is decreased in $Shank3B^{-/-}$ mice (red trace) as measured by extracellular field recordings. Inset, example traces for $Shank3B^{-/-}$ (KO) and wild-type (WT). **b**, mEPSC example traces from wild-type and $Shank3B^{-/-}$ MSNs recorded with whole-cell voltage clamp. **c, d**, Reduced mEPSC frequency (**c**) and amplitude (**d**) in $Shank3B^{-/-}$ MSNs when compared to wild-type. **e**, PPR is unaltered in $Shank3B^{-/-}$ MSNs. **$P < 0.01$, ***$P < 0.001$; two-way repeated measures ANOVA, with Bonferroni post hoc test for **a**; two-tailed t-test for **c**, **d**; all data presented as means ± s.e.m. For field recordings, $n = 13$ slices from four mice per group; for mEPSCs, $n = 29$ MSNs from wild-type mice, $n = 32$ MSNs from $Shank3B^{-/-}$ mice.

We next performed whole-cell voltage clamp recordings of α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor-miniature excitatory postsynaptic currents (AMPAR-mEPSCs) in dorsolateral striatal MSNs. We found that the frequency of mEPSCs was significantly reduced in $Shank3B^{-/-}$ MSNs (Fig. 5b, c), indicating a reduction in the number of functional synapses in $Shank3B^{-/-}$ MSNs because we did not observe defects on presynaptic function by measuring PPR (Fig. 5e). We also found a significant reduction of peak mEPSC amplitude in $Shank3B^{-/-}$ MSNs (Fig. 5b, d), indicating a reduction in the postsynaptic response from the available synapses. We did not observe significant differences in N-methyl-D-aspartate receptor (NMDA)/AMPA receptor-mediated current ratio in $Shank3B^{-/-}$ neurons (Supplementary Fig. 12). Finally, similar defects in mEPSC frequency and amplitude were observed in $Shank3B^{-/-}$ and wild-type littermate mice obtained from heterozygous matings (Supplementary Fig. 13). Together, these data demonstrate a critical role for SHANK3 in postsynaptic function in cortico-striatal circuitry.

To assess if the defects arising from Shank3 dysfunction were specific to striatal circuitry or due to a more broad CNS perturbation, we performed a Morris water maze task for hippocampal-dependent learning and memory. We found that $Shank3B^{-/-}$ mice performed at the same levels as controls in both learning and probe trials (Supplementary Fig. 14a–c). Reversal learning and probe trials again demonstrated similar levels of performance between $Shank3B^{-/-}$ mice and controls (Supplementary Fig. 14d–f). Concomitantly, we performed electrophysiological recordings from the hippocampal CA1 sub-region and found no obvious difference in field recordings of population spikes or PPR between genotypes (Supplementary Fig. 15a–c). In addition, we found no significant differences in mEPSC frequency or mEPSC amplitude (Supplementary Fig. 15d–f). These data suggest that the observed behavioural and synaptic defects are specific to discrete brain regions and are not part of an overall CNS dysfunction.

Discussion

Despite recent advances in the understanding of autism spectrum disorder genetics, the underlying neurobiological substrates and neural circuits involved in these disorders remain largely unknown. The Shank3 gene has become the focus of substantial interest, with an increasing body of evidences suggesting Shank3 as the causative gene of the major neurological symptoms in the 22q13 deletion syndrome[9,11–13,24]. Our present study with Shank3 mutant mice not only sheds light on a critical in vivo role for SHANK3 in striatal glutamatergic synaptic structure and function, but also demonstrates causality between a disruption in this gene and the development of autistic-like behaviours in mice.

In this study, we generated two mutant alleles for the Shank3 gene. These two lines of mice showed different levels of severity in synaptic defects and phenotypes. In humans, multiple mutations/variants of Shank3 gene have been identified to coalesce at the ankyrin repeats and downstream of PDZ domain[9,13]. Our data indicate that disruptions of different locations of the Shank3 gene can lead to varying degrees of functional defects, which may in part contribute to phenotypic heterogeneity in Shank3-related ASDs. We should note that, in clinical conditions, the 22q13 deletions and the autism-associated Shank3 mutations are heterozygous, whereas in our current study, we used homozygous mutant mice to get a clear understanding of the physiological role of the Shank3 gene and the underlying functional consequences of its disruption. Further studies will be needed to elucidate potential functional deficits resulting from Shank3 haploinsufficiency in $Shank3B^{-/-}$ mice.

PSD-95–SAPAP–SHANK proteins form a key postsynaptic scaffold at glutamatergic synapses which interacts with many synaptic proteins, including the neurexin–neuroligin complex[25]. In addition to Shank3 (ref. 9), it is worth noting that Shank2 (refs 6, 8), SAPAP2 (ref. 6), neurexin-1 (ref. 26) and neuroligin-3 and -4 (ref. 27) have all been implicated in human ASDs. Therefore, the dysfunction of neurexin–neuroligin–PSD-95–SAPAP–SHANK complex could underlie a common synaptic mechanism for a subset of ASDs.

The precise circuitry defects involved in autistic behaviours are poorly understood. Neuroimaging studies provide evidence that caudate and frontal-striatal circuitries are dysfunctional areas in ASD[28–30]. Cortico-striatal circuitry dysfunction has also been strongly implicated in repetitive/compulsive behaviours in obsessive-compulsive disorder (OCD)[31–33]. We previously found that deletion of SAPAP3, which directly interacts with SHANK3 and is highly expressed in the striatum, leads to cortico-striatal circuitry dysfunction and OCD-like behaviours including repetitive/compulsive grooming in mice[31]. Repetitive behaviours are also often seen in autistic patients and in some mouse models of ASDs[34–36]. SHANK3 is the most abundant SHANK family member expressed in the striatum and $Shank3B^{-/-}$ mice exhibit excessive/repetitive grooming leading to skin lesions. Our data support the hypothesis that repetitive behaviours in OCD and ASD may share a common circuitry mechanism.

The regulation of social behaviours and social interaction is thought to be controlled by several brain regions and circuits[37]. Similarly, genetic makeup is thought to have a key role in the phenotypical manifestation of social behaviours[38]. The robust social interaction deficits in Shank3B mutant mice demonstrate a casual role for the disruption of this gene in the genesis of social dysfunction and provide a valuable experimental system for future genetic dissection of the neuronal basis of social behaviour.

METHODS SUMMARY

Behavioural analysis. Young adult mice 5–6-weeks old were used for all behavioural analyses except lesion scores which were performed in 4–5-month-old mice. All experiments were done blind to genotypes. All experimental procedures were reviewed and approved by the Duke University Institutional Animal Care and Use Committee and the MIT Committee on Animal Cares.

Statistical analysis. Analyses were performed using Prism (GraphPad Software) and MATLAB (MathWorks). Details on particular tests used are described in the main text and in the methods section; a summary of statistical analysis for the behavioural data are presented in Supplementary Table 2.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. American Psychiatric Association Task Force on DSM-IV. *Diagnostic and statistical manual of mental disorders: DSM-IV-TR* (American Psychiatric Association, 2000).
2. Rosenberg, R. E. *et al.* Characteristics and concordance of autism spectrum disorders among 277 twin pairs. *Arch. Pediatr. Adolesc. Med.* **163,** 907–914 (2009).
3. Abrahams, B. S. & Geschwind, D. H. Advances in autism genetics: on the threshold of a new neurobiology. *Nature Rev. Genet.* **9,** 341–355 (2008).
4. Bourgeron, T. A synaptic trek to autism. *Curr. Opin. Neurobiol.* **19,** 231–234 (2009).
5. Zoghbi, H. Y. Postnatal neurodevelopmental disorders: meeting at the synapse? *Science* **302,** 826–830 (2003).
6. Pinto, D. *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466,** 368–372 (2010).
7. Tabuchi, K. *et al.* A neuroligin-3 mutation implicated in autism increases inhibitory synaptic transmission in mice. *Science* **318,** 71–76 (2007).
8. Berkel, S. *et al.* Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation. *Nature Genet.* **42,** 489–491 (2010).
9. Durand, C. M. *et al.* Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. *Nature Genet.* **39,** 25–27 (2006).
10. Prasad, C. *et al.* Genetic evaluation of pervasive developmental disorders: the terminal 22q13 deletion syndrome may represent a recognizable phenotype. *Clin. Genet.* **57,** 103–109 (2000).
11. Wilson, H. L. *et al.* Molecular characterisation of the 22q13 deletion syndrome supports the role of haploinsufficiency of SHANK3/PROSAP2 in the major neurological symptoms. *J. Med. Genet.* **40,** 575–584 (2003).
12. Moessner, R. *et al.* Contribution of SHANK3 mutations to autism spectrum disorder. *Am. J. Hum. Genet.* **81,** 1289–1297 (2007).
13. Gauthier, J. *et al.* Novel de novo SHANK3 mutation in autistic patients. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **150B,** 421–424 (2009).
14. Kim, E. *et al.* GKAP, a novel synaptic protein that interacts with the guanylate kinase-like domain of the PSD-95/SAP90 family of channel clustering molecules. *J. Cell Biol.* **136,** 669–678 (1997).

15. Takeuchi, M. *et al.* SAPAPs. A family of PSD-95/SAP90-associated proteins localized at postsynaptic density. *J. Biol. Chem.* **272,** 11943–11951 (1997).
16. Zoghbi, H. Y. & Warren, S. T. Neurogenetics: advancing the ''next-generation'' of brain research. *Neuron* **68,** 165–173 (2010).
17. Moy, S. S. *et al.* Sociability and preference for social novelty in five inbred strains: an approach to assess autistic-like behavior in mice. *Genes Brain Behav.* **3,** 287–302 (2004).
18. Hung, A. Y. *et al.* Smaller dendritic spines, weaker synaptic transmission, but enhanced spatial learning in mice lacking Shank1. *J. Neurosci.* **28,** 1697–1708 (2008).
19. Redcay, E. & Courchesne, E. When is the brain enlarged in autism? A meta-analysis of all brain size reports. *Biol. Psychiatry* **58,** 1–9 (2005).
20. Langen, M. *et al.* Changes in the developmental trajectories of striatum in autism. *Biol. Psychiatry* **66,** 327–333 (2009).
21. Hollander, E. *et al.* Striatal volume on magnetic resonance imaging and repetitive behaviors in autism. *Biol. Psychiatry* **58,** 226–232 (2005).
22. Bourgeron, T. A synaptic trek to autism. *Curr. Opin. Neurobiol.* **19,** 231–234 (2009).
23. Kwon, C. H. *et al.* Pten regulates neuronal arborization and social interaction in mice. *Neuron* **50,** 377–388 (2006).
24. Bonaglia, M. C. *et al.* Identification of a recurrent breakpoint within the *SHANK3* gene in the 22q13.3 deletion syndrome. *J. Med. Genet.* **43,** 822–828 (2006).
25. Irie, M. *et al.* Binding of neuroligins to PSD-95. *Science* **277,** 1511–1515 (1997).
26. Kim, H. G. *et al.* Disruption of neurexin 1 associated with autism spectrum disorder. *Am. J. Hum. Genet.* **82,** 199–207 (2008).
27. Jamain, S. *et al.* Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nature Genet.* **34,** 27–29 (2003).
28. Silk, T. J. *et al.* Visuospatial processing and the function of prefrontal-parietal networks in autism spectrum disorders: a functional MRI study. *Am. J. Psychiatry* **163,** 1440–1443 (2006).
29. Horwitz, B., Rumsey, J. M., Grady, C. L. & Rapoport, S. I. The cerebral metabolic landscape in autism. Intercorrelations of regional glucose utilization. *Arch. Neurol.* **45,** 749–755 (1988).
30. Sears, L. L. *et al.* An MRI study of the basal ganglia in autism. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **23,** 613–624 (1999).
31. Welch, J. M. *et al.* Cortico-striatal synaptic defects and OCD-like behaviours in *Sapap3*-mutant mice. *Nature* **448,** 894–900 (2007).
32. Shmelkov, S. V. *et al.* Slitrk5 deficiency impairs corticostriatal circuitry and leads to obsessive-compulsive-like behaviors in mice. *Nature Med.* **16,** 598–602 (2010).
33. Graybiel, A. M. Habits, rituals, and the evaluative brain. *Annu. Rev. Neurosci.* **31,** 359–387 (2008).
34. McFarlane, H. G. *et al.* Autism-like behavioral phenotypes in BTBR T+tf/J mice. *Genes Brain Behav.* **7,** 152–163 (2008).
35. Blundell, J. *et al.* Neuroligin-1 deletion results in impaired spatial memory and increased repetitive behavior. *J. Neurosci.* **30,** 2115–2129 (2010).
36. Etherton, M. R., Blaiss, C. A., Powell, C. M. & Sudhof, T. C. Mouse neurexin-1α deletion causes correlated electrophysiological and behavioral changes consistent with cognitive impairments. *Proc. Natl Acad. Sci. USA* **106,** 17998–18003 (2009).
37. Insel, T. R. & Fernald, R. D. How the brain processes social information: searching for the social brain. *Annu. Rev. Neurosci.* **27,** 697–722 (2004).
38. Ebstein, R. P., Israel, S., Chew, S. H., Zhong, S. & Knafo, A. Genetics of human social behavior. *Neuron* **65,** 831–844 (2010).

METHODS

Mice. *Shank3* mutant mice were generated by homologous recombination in R1 embryonic stem cells and implanted in C57 blastocysts using standard procedures. One targeting vector (Shank3A) was designed to replace exon 4–7 (containing the ankyrin repeat domains) and another vector (Shank3B) was designed to replace exon 13–16 (containing the PDZ domain) of the *Shank3* gene with a NEO cassette. Genotypes were determined by PCR of mouse tail DNA, using: for Shank3A, primer F1a (GGTTGAGGATGAGCAAGCTAG) and R1a (GGGAC ATAAGTGAAGGTTAGG) for the wild-type allele (318 base pairs), and F1a and R2 (TCAGGGTTATTGTCTCATGAGC; in the neo cassette) for the mutant allele (361 base pairs); for Shank3B, primer F1b (GAGCTCTACTCCCTT AGGACTT) and R1b (TCCCCCTTTCACTGGACACCC) for the wild-type allele (316 base pairs), and F1b and R2 (TCAGGGTTATTGTCTCATGAGC; in the neo cassette) for the mutant allele (360 base pairs). The NEO cassette was not removed.

Chimaeric mice were crossed to C57 females (Jackson Labs). Initially, F1 hybrids from heterozygous × heterozygous matings were generated. However, homozygous knockouts mice from this type of mating are smaller than their wild-type littermates, presumably due to an inadequate competition for resources during early postnatal days leading to different developmental trajectories. We postulated that this size difference would influence our behavioural tests. To alleviate this confound, heterozygous animals were crossed in direct brother-sister matings for five generations from which we derived F5 isogenic hybrids in a mixed background. These isogenic animals were then used to generate time-mated homozygous × homozygous breeding pairs to obtain wild-type and mutant animals used in the experiments. F5 *Shank3A* and F5 *Shank3B* knockouts from these matings are reared to weaning age with weights similar to those from F5 control animals.

Animals were housed at a constant 23 °C in a 12 h light/dark cycle (lights off at 19:00), with food and water available *ad libitum*. Mice were housed 3–5 by genotype per cage with the exception of the animals individually housed for grooming measurements. Only aged-matched male mice were used for behavioural experiments, all other tests included age-matched males and females in proportional contribution across groups. Unless otherwise noted, all tests were conducted with naive cohorts of mice. All experimental procedures were reviewed and approved by the Duke University Institutional Animal Care and Use Committee and the MIT Committee on Animal Cares..

Grooming behaviour[31]. Young adult male mice 5–6-week-old were used for analysis of grooming behaviour. Habituated, individually housed animals were video-taped for 24 h under 700 lx (day, 12 h) and ~2 lx (red light at night, 12 h) illumination. Grooming behaviours were coded from 19:00–21:00 h (that is, 2 h beginning at the initiation of the dark cycle); this segment was analysed using Noldus Observer software and the total amount of time in the 2-h segment spent grooming was determined. Grooming included all sequences of face-wiping, scratching/rubbing of head and ears, and full-body grooming. The observer was blinded to genotype during the scoring of the videotapes.

PSD preparation and western blot. PSD fractions of the striatum were prepared as previously described[31], separated on SDS–PAGE and probed with specific antibodies. The relative amount of β-actin was used as loading control. Antibodies used in these experiments include rabbit antibodies against PSD-93 (gift from D. Bredt) and Shank3 (gift from T. Boeckers). The antibody for SAPAP3 has been previously described[39]. Commercial antibodies used include monoclonal antibodies against NR1 (Transduction Laboratories), NR2B (Millipore), CaMKII (Transduction Laboratories), NR2A (Millipore), and β-actin (Sigma), as well as polyclonal antibodies against GluR1 (Abcam), Homer (Chemicon), GluR2 (Abcam), neuroligin-3 (Synaptic Systems) and PSD-95 (Abcam).

***In situ* hybridization.** mRNA *in situ* hybridization was performed as described elsewhere[39]. Briefly, reactions were performed with 20 μm cryosections from freshly frozen 5-week-old brain mouse tissue using digoxigenin (DIG)-labelled riboprobes against mouse *Shank1* cDNA (NM_001034115; base pairs 4107–4924), *Shank2* cDNA (NM_001081370; base pairs 2063–2876) and *Shank3* (NM_021423; base pairs 3159–3959). The complementary DNAs used were all verified by sequencing compared to the following sequences GenBank accession numbers: (*Shank1*: NM_001034115), (*Shank2*: NM_001081370) and (*Shank3*: NM_021423). The hybridization signal was detected using an alkaline phosphatase (AP)-conjugated anti-DIG antibody (Roche) and developed using 5-bromo-4-cloro-indolylphosphate/nitroblue tetrazolium (BCIP/NBT; Roche).

Motor and anxiety-like behaviours[31]. Zero maze: an elevated zero maze was indirectly illuminated at 100 lx. Testing commenced with an animal being introduced into a closed area of the maze. Behaviour was video-taped for 5 min and subsequently scored by a trained observer using Noldus Observer software. Anxiety-like behaviour was deduced based upon the percent time spent in the open areas. The observer was blinded to genotype. The animals used in the zero

maze test, both *Shank3A*$^{-/-}$, *Shank3B*$^{-/-}$ and respective controls were previously tested in the open field test with a 2-day period in between tasks.

Open field: spontaneous locomotor activity was evaluated over 30 min in an automated Omnitech Digiscan apparatus (AccuScan Instruments) as described[31]. Locomotor activity was assessed as total distance travelled (m). Anxiety-like behaviour was defined by number of rearings and time spent in the centre as compared to time spent in the perimeter (thigmotaxis) of the open field.

Dark-light emergence test: mice were habituated in an adjacent room to low light conditions (~40 lx) and the test room was initially under similar illumination. Testing was conducted in a two-chambered test apparatus (Med Associates), with one side draped in black cloth (that is, dark-chamber) and the other illuminated at ~1,000 lx (that is, light-chamber) with a high intensity house light and overhead fluorescent lamps. Upon placing the mice into the dark chamber, the light chamber was illuminated and the door between the two chambers was opened. The mice were allowed to freely explore the apparatus for 5 min. The latency to emerge from the darkened into the lighted chamber and the percentage of time spent in the illuminated chamber were used as indices of anxiety-like behaviours.

Social interaction paradigm. Three-chamber social test: sociability and response to social novelty test was performed as previously described[17] with minor modifications. Briefly, 5–6-week-old male animals were used across all tests. Target subjects (Stranger 1 and Stranger 2) were 5–6-week-old males habituated to being placed inside wire cages for 5 days before beginning of testing. Test mice were habituated to the testing room for at least 45 min before the start of behavioural tasks. The social test apparatus consisted of a transparent acrylic box with removable floor and partitions dividing the box into three chambers. Here, the middle chamber (20 cm × 17.5 cm) is half the width of Chamber 1 (20 cm × 35 cm) and Chamber 2 (20 cm × 35 cm) with the overall dimensions of the box being 60 cm (length) × 35 cm (width) with 5 cm openings between each chamber which can be closed or open with a lever operated door. The wire cages used to contain the stranger mice were cylindrical, 11 cm in height, a bottom diameter of 10.5 cm with the bars spaced 1 cm apart (Galaxy Cup, Spectrum Diversified Designs). An inverted transparent cup was placed on the top of the cage to prevent the test mice from climbing on the top of the wire cage.

For the sociability test, the test animal was introduced to the middle chamber and left to habituate for 5 min, after which an unfamiliar mouse (Stranger 1) is introduced into a wire cage in one of the side-chambers and an empty wire cage on the other side-chamber. The dividers are then raised and the test animal is allowed to freely explore all three chambers over a 5 min session. Following the 5 min session, the animal remains in the chamber for an extra 5 min (post-test) to better acquire the identification cues from Stranger 1 animal. Following this, a novel stranger mouse (Stranger 2) is inserted in the wire cage previously empty and again the test animal is left to explore for a 5 min session. Time spent in each chamber, time spent in close proximity and heat maps were calculated using the automated software Noldus Ethovison. The release of the animals and relative position of social and inanimate targets was counterbalanced. However, for each individual test animal the location of Stranger 1 was maintained during Stranger 1 – E and Stranger 1 – Stranger 2 testing of the social behaviour.

Dyadic social interaction: animals were acclimatized to the test room for at least 1 h before the experiment. Target mice were wild-type and *Shank3B*$^{-/-}$ of 6 weeks of age. Stimulus mice were conspecific age-matched wild-type mice socially naive to the target mice. At least 3 h before the beginning of the test, stimulus mice were given identifiable markings on the tails using a black marker pen. A pair of target and stimulus mice were introduced in a transparent Plexiglas arena (40 cm × 40 cm × 30 cm) covered with fresh bedding and the session recorded for 10 min. Quantification of social behaviours was performed using Noldus Observer software by a researcher blinded to the genotype of the target animals. Quantifications included: reciprocal social interaction, as determined by any sequence or combination of sequences involving close huddling, sniffing (for example, nose-to-nose, anogenital sniffing) or allogrooming by the target and stimulus mouse; the frequency of nose-to-nose sniffing; and the frequency of anogenital sniffing initiated by the target animal towards the stimulus mouse. Statistical tests were performed using unpaired two-tailed *t*-test.

Rotarod. Motor coordination was assessed in an accelerating rotarod test (4–40 r.p.m.). Briefly, animals were introduced in the apparatus (Med Associates) and the latency to fall was determined. Animals were tested for three trials in a single day with an inter-trial interval of 10 min.

Morris water maze. Morris water maze testing was conducted as describe elsewhere[40] with minor modifications. Male mice (4–5-weeks old) selected for the test were individually handled daily for 5 days before beginning the experiment. Testing pool was 120 cm in diameter and the platform 8 cm in diameter. The platform was submerged 1 cm below the water surface. Pool water was maintained at 23.0 ± 0.5 °C and made opaque by mixing-in white non-toxic tempera paint. During training, 90 s duration trials were used, if the animals did not find

the platform within 90 s the experimenter guided the animal to the platform. After reaching the platform the animals were left for 15 s on top of the platform before being removed. Trials were administered for 5 days with four trials per animal per day with the platform located in the south-west quadrant. On the sixth day a 60 s probe trial was performed. On the seventh day, the reversal training commenced with the platform in the north-east quadrant, and proceeded as described above. The experimenter followed the animals' progress using tracking software outside of the testing room. Tracking and analysis were performed using the Noldus Ethovison software.

Golgi staining and Sholl analysis. All brains and collected sections were coded in order to blind the experimenter of the genotype until after all data was collected and analysed. Brains from 5-week-old, gender-matched $Shank3B^{-/-}$ and control mice were prepared using standard Golgi–Cox impregnation technique using the FD Rapid GolgiStain Kit (NeuroTechnologies). Serial coronal sections of 100 μm were collected from controls and $Shank3B$ mutant animals. A total of 12 cells per animal were traced across the dorsal striatum as to sample representatively from this structure for a final number of 36 cells per genotype. For each animal, sections were selected to be between rostral-caudal bregma 1.18 mm and 0.86 mm. Criteria to identify medium spiny neurons were, (1) presence within the caudate putamen; (2) full impregnation of the neuron along the entire length of the dendritic arborization; (3) relative non-overlap with surrounding neurons and isolation from astrocytes and blood vessels and (4) morphologically, by the presence of high number of spines and relatively short neuronal arborizations as characteristics of MSNs. For each selected neuron the entire neuronal arbor was reconstructed under a ×100 oil lens in a motorized microscope with a digital CCD camera connected to a computer running Neurolucida Software (MBF Bioscience). The three-dimensional analysis of the reconstructed neurons was performed using NeuroExplorer software (MBF Bioscience) and data from branch length, number of branches and neuronal complexity was measured and analysed in Prism (Graph Pad). Two-way repeated measures ANOVA was used for Sholl analysis. Statistical significance was accepted when *$P < 0.05$, **$P < 0.01$ and ***$P < 0.0001$.

Cortico-striatal electrophysiology. Brain slice preparation for extracellular field recording: acute brain slices were prepared from 6–7-week-old mice. Slices were prepared from one WT and one KO pair each day and the experimenter was blinded to the genotype. The mice were deeply anesthetized by intra-peritoneal injection of avertin and then transcardially perfused with carbogenated (95% O_2, 5% CO_2) ice-cold protective cutting artificial cerebrospinal fluid (aCSF) with the composition (in mM): 119 glycerol, 2.5 KCl, 1.25 NaH_2PO_4, 26 $NaHCO_3$, 25 glucose, 2 thiourea, 5 L-ascorbic acid, 3 Na-pyruvate, 0.5 $CaCl_2$.$4H_2O$, 10 $MgSO_4$.$7H_2O$. Mice were then decapitated and the brains were removed into ice-cold cutting solution for an additional 1 min. The brains were then rapidly blocked for coronal sectioning at 300-μm thickness on a VF200 model compresstome (Precisionary Instruments) using either a sapphire or zirconium ceramic injector style blade. Slices containing the dorsal striatum were initially recovered for 30 min at room temperature (23–25 °C) in a carbogenated protective recovery aCSF (same composition as the cutting aCSF except that glycerol was replaced with N-methyl-D-glucamine (NMDG)-Cl as a substitute for NaCl to prevent initial excitotoxic swelling during re-warming). After this initial 30 min period the slices were transferred into a holding chamber containing carbogenated normal aCSF of the composition (in mM): 119 NaCl, 2.5 KCl, 1.25 NaH_2PO_4, 26 $NaHCO_3$, 12.5 glucose, 2 $CaCl_2$.$4H_2O$, 1 $MgSO_4$.$7H_2O$. The holding aCSF was supplemented with (in mM): 2 thiourea, 1 L-ascorbic acid, 3 Na-pyruvate to improve slice health and longevity, and slices were stored for 1–6 h before transfer to the recording chamber for use. The osmolarity of all solutions was measured at 300–310 mOsm and the pH was maintained at ~7.3 after equilibration under constant carbogenation.

Supplementary Fig. 11 shows summary data for corticostriatal field recordings from acute coronal brain slices of Shank3A mutant versus WT mice. The method of slice preparation differed significantly in these earlier experiments. Mice were transcardially perfused with carbogenated ice-cold protective aCSF with the composition (in mM): 185 sucrose, 2.5 KCl, 1.25 NaH_2PO_4, 26 $NaHCO_3$, 25 glucose, 0.5 $CaCl_2$.$4H_2O$, 4 $MgSO_4$.$7H_2O$ (pH 7.3, 300–310 mOsm) without supplementation of antioxidants. Slices were immediately transferred into a holding chamber containing carbogenated normal aCSF of the composition (in mM): 119 NaCl, 2.5 KCl, 1.25 NaH_2PO_4, 26 $NaHCO_3$, 12.5 glucose, 2 $CaCl_2$.$4H_2O$, 1 $MgSO_4$.$7H_2O$ (pH 7.3, 300–310 mOsm) without supplementation of antioxidants, and slices were stored for 1–4 h before transfer to the recording chamber. The absence of the initial 30 min recovery period in 'protective' aCSF in addition to the absence of antioxidant supplementation in the cutting aCSF and in the aCSF in the holding chamber results in more rapid deterioration of slice health and smaller evoked population spike amplitudes on average, indicating reduced overall slice viability compared to slices prepared with a 30 min NMDG aCSF recovery protocol described above. However, WT and KO brain slices were always subjected to identical procedures on any given day of recording and the procedures were always standardized for each discrete experimental data set so that these factors would not introduce any potential confounds.

Extracellular field recording. A platinum iridium concentric bipolar stimulating electrode (CBAPC75, 25 μm inner pole diameter; FHC) was placed on the inner border of the corpus callosum between the cortex and dorsolateral striatum. This electrode position was chosen to predominantly activate cortical axons within the corpus callosum which heavily converge upon striatal MSNs to form excitatory corticostriatal synaptic connections. Although there is ample evidence on which to base our assertion that stimulation of the corpus callosum predominantly results in activation of cortical axons[41,42], we are unable to exclude the possibility of a relatively smaller contribution arising from activation of thalamostriatal axons that have distal terminals in dorsolateral striatum nearby to the stimulated region. Thus, although we refer to our measurements as primarily reflecting corticostriatal transmission, our measurements are not 'pure' corticostriatal responses. Borosilicate glass recording electrodes filled with 2 M NaCl were placed in the dorsolateral striatum approximately 400–450 μm away from the stimulating electrode. Corticostriatal field population spikes were evoked with 0.15 ms step depolarizations at 0.5 mA intensity at a frequency of 0.05–0.1 Hz. Paired pulses were evoked with a 100 ms inter-stimulus interval. Baseline responses were monitored to ensure stable population spike amplitude for a minimum of 5 min. Input-output functions were then determined for the negative peak 1 (NP1; presynaptic fibre volley) and population spike amplitude by three consecutive rounds of stimulation from 0–1.0 mA in 0.1 mA increments. All recordings were performed at room temperature and acquired using pCLAMP 10 software (Axon Instruments/Molecular Devices). Data analysis was performed blind to genotype in Clamp fit (Axon Instruments/Molecular Devices). Population spike amplitude was measured as the average of the early peak positivity to the peak negativity and from the peak negativity to the late peak positivity. This standard method takes into account the fact that the downward population spike is superimposed on an upward field excitatory postsynaptic potential (fEPSP). Paired pulse ratio (PPR) was calculated as the ratio of the 2nd population spike amplitude to the 1st population spike amplitude for responses to paired pulse stimulation at 0.5 mA fixed intensity with a 100 ms inter-stimulus interval for the pair.

Extracellular field recordings and whole-cell mEPSC recordings in the hippocampal CA1 region were conducted in 300 μm thick acute brain slices from 6–9-week-old WT and Shank3B mutant mice. For measurement of hippocampal CA1 population spikes, a concentric bipolar stimulating electrode was placed in the stratum radiatum to stimulate the Schaffer collateral pathway, and a borosilicate glass recording electrode (~2–3 MΩ) filled with recording aCSF was placed in the CA1 pyramidal cell layer approximately 400 μm from the stimulation site. The recording electrode was placed at the depth in the slice that gave the largest population spike amplitude, and a stable baseline was established for <10 min. Input-output recordings were conducted by increasing the stimulation intensity from 0 to 160 μA in 20 μA increments. Three successive rounds were collected and values at each intensity represent the average of the three measurements. CA1 population spike amplitude was quantified exactly as described previously for cortico-striatal population spikes. For CA1 pyramidal neuron whole-cell recordings, pyramidal neurons in CA1 were identified under infrared-differential interference contrast (IR-DIC) visualization. Cells were patched with a Caesium-gluconate-based internal solution containing (in mM): 110 Caesium-gluconate, 15 KCl, 4 NaCl, 5 TEA-Cl, 20 HEPES, 0.2 EGTA, 5 lidocaine N-ethyl chloride, 4 ATP magnesium salt, and 0.3 GTP sodium salt. The pH was adjusted to 7.25 with D-gluconic acid and osmolarity was adjusted to 290–300 mOsm with sucrose as necessary. The recording aCSF contained 1 μM TTX, 100 μM picrotoxin, 5 μM CGP55845, and 50 μM D-APV to isolate pure AMPAR-mediated mEPSCs. CA1 neurons were voltage-clamped at −80 mV to amplify the smallest spontaneous miniature synaptic events that might otherwise escape detection. Criteria for acceptance were uncompensated stable $R_a < 25$ MΩ and holding current < -300 pA. mEPSCs were detected using MiniAnalysis software (Synaptosoft) as described for striatal MSNs. All recordings were carried out at room temperature (23–25 °C). Slices were prepared in a 20–30 degree off-horizontal cutting angle (optimal for CA1 region) from one WT and one KO pair each day and the experimenter was blind to the genotypes of the animals.

Striatal slice preparation for whole-cell recording. Mice 5–6-week-old were used for all whole-cell electrophysiology procedures by an experimentalist blinded to genotype. Acute coronal striatal slices were prepared as follows. Briefly, mice were anesthetized with Avertin solution (20 mg/ml, 0.5 mg/g body weight) and perfused through the heart with a small volume (about 20 ml) of ice-cold and oxygenated (95% O_2, 5% CO_2) cutting solution containing (mM): 105 NMDG, 105 HCl, 2.5 KCl, 1.2 NaH_2PO_4, 26 $NaHCO_3$, 25 Glucose, 10 $MgSO_4$, 0.5 $CaCl_2$, 5 L-Ascorbic Acid, 3 Sodium Pyruvate, 2 Thiourea (pH 7.4, with osmolarity of 295–305 mOsm). The brains were rapidly removed and placed in ice-cold and oxygenated cutting solution. The coronal slices (300 μm) were prepared

using a slicer (Vibratome 1000 Plus, Leica Microsystems, USA) and then transferred to an incubation chamber (BSK4, Scientific System Design Inc., USA) at 32 °C with carbogenated cutting solution, which was gradually replaced with aCSF in 30 min through a peristaltic pump (Rainin, RP-1) allowing a precise regulation of flowing rates. The slices were then kept in the aCSF that contained (mM): 119 NaCl, 2.3 KCl, 1.0 NaH$_2$PO$_4$, 26 NaHCO$_3$, 11 Glucose, 1.3 MgSO$_4$, 2.5 CaCl$_2$ (pH was adjusted to 7.4 with HCl, with osmolarity of 295–305 mOsm) at room temperature for at least 30 min.

Whole-cell patch-clamp. The slice was placed in a recording chamber (RC-27L, Warner Instruments) and constantly perfused with oxygenated aCSF at 24 °C (TC-324B, Warner Instruments) at a rate of 1.5–2.0 ml min^{-1}. The striatum and individual MSNs were visualized and identified with a microscope equipped with IR-DIC optics (BX-51WI, Olympus) by location, shape and size (ovoid cell body with major axis of 10 to 14 μm). Two additional measures were used to distinguish them from similar sized GABAergic interneurons. First, GABAergic interneurons show smaller membrane capacitance (C_m) and membrane time constant (τ_m) (at least two times less) when compared to that of MSNs. In the case of recordings done with Cs$^+$ internal, these membrane properties were measured immediately after membrane rupture when the Cs$^+$ internal has not been dialysed and taken effect yet. Second, AMPA receptor–mediated mEPSCs showed much faster kinetics (including both rise time and decay time constant, τ decay) in GABAergic interneurons. Whole-cell patch-clamp recordings were obtained from MSNs using recording pipettes (King Precision Glass, glass type 8250) pulled in a horizontal pipette puller (P-87, Sutter Instruments) to a resistance of 3–4 MΩ, when filled with the internal solution containing (in mM): 107 CsMeSO$_3$, 10 CsCl, 3.7 NaCl, 5 TEA-Cl, 20 HEPES, 0.2 EGTA, 5 lidocaine N-ethyl chloride, 4 ATP magnesium salt, and 0.3 GTP sodium salt. pH was adjusted to 7.3 with KOH and osmolarity was adjusted to 298–300 mOsm with 15 mM K$_2$SO4.

To record AMPA receptor-miniature excitatory postsynaptic currents (mEPSCs), the cells were held in voltage clamp at −70 mV in the presence of 50 μM APV (DL-2-amino-5-phosphono-valeric acid), 25 μM BMR (1(S),9(R)-(−)-bicuculline methbromide), 10 μM D-serine and 1 μM TTX (all from Tocris). The miniature events were not recorded until 5 min after entering whole cell patch clamp recording mode to allow the dialysis of Cs$^+$ internal solution for a relatively complete block of the potassium channels in the MSNs. The mEPSCs were detected and analysed with MiniAnalysis (Synaptosoft).

For paired-pulse stimulation experiments, AMPAR mediated excitatory postsynaptic currents (EPSCs) were evoked by a local concentric bipolar stimulating electrode (CBARC75, FHC) that was placed at the inner edge of corpus callosum within the dorso-lateral region of the striatum. Recordings were made in the presence of picrotoxin (100 μM) and APV (50 μM) to block activation of GABA$_A$ receptors and NMDA receptors. Stimulation was current-controlled (ISO-Flex, A.M.P.I.). The stimulus intensity was set at a level that could evoke 300–400 pA AMPAR-mediated response for all the cells measured and delivered with an inter-stimulus interval of 50 ms. The paired-pulse measurements were obtained for 15–20 consecutive traces and only those traces with stable evoked first current response were used for data analysis. The PPR was calculated with the peak current response to the second pulse divided by that of the first response.

NMDAR- and AMPAR- mediated synaptic current ratio (NMDA/AMPA ratio) was recorded in the presence of picrotoxin at holding potentials of +40 mV and −70 mV, respectively. The NMDA/AMPA ratios were measured according to previously described methods[43]. Briefly, the stimulus intensity was set at a level that could evoke 300–400 pA AMPAR-mediated response with a holding potential at −70 mV. Each evoked response was repeated for 15–20 times with an inter-stimulus interval of 20 s for all the cells measured. The time point of the peak current at –70 mV, considered to be fully mediated by AMPARs, was used to establish the time window for measuring the AMPA peak at +40 mV. The decay to baseline of the AMPA current at –70 mV was used to select a time window for measurement of the NMDA current; a 10-ms measurement window beginning 40 ms after the stimulus artefact was used. This current amplitude at this point was designated as the NMDAR mediated synaptic current response. (I_{NMDA} at +40 mV/I_{AMPA} at –70 mV) was taken as the NMDA/AMPA ratio.

Data acquisition and analysis. A Multiclamp 700B amplifier (Molecular Devices Corporation) and digidata 1440A were used to acquire whole cell signals. The signals were acquired at 20 kHz and filtered at 2 kHz. The series-resistance was <20 MΩ. Values are expressed as means ± s.e.m. Data were tested for significance using either an unpaired t-test or a two-way repeated measures ANOVA.

Cell filling. Mice were assigned a code previous to dissection, as to maintain a blinded genotype across all procedures, including dissection, cell filling, imaging and quantification. Mice were deeply anesthetized with an overdose of isofluorane and transcardially perfused with PBS (pH 7.4) followed by ice-cold 4% paraformaldehyde/PBS (PFA) (pH 7.4). The brains were removed and post-fixed

overnight in PFA 4%. After post-fixation, the brain was sliced at 200-μm thickness coronal sections in a vibratome and kept in PBS at 4 °C. For cell filling injections, selected brain slices immersed in PBS were mounted in a tissue stage. Dorsal striatal medium cells were targeted with post hoc confirmation of being medium spiny neurons (morphology and spine density). Using a micromanipulator, micropipettes loaded with Lucifer Yellow dye (Sigma L-0259, 8% solution in 0.05 M Tris buffer, pH 7.4) were used to impale the cell body. A micropipette containing a solution of 0.1 M LiCl was used to deliver the dye with a continuous 10 nA current for 5 min. Following cell filling, a post-staining was used to amplify the fluorescent signal. Briefly, sections were transferred to blocking solution (5% sucrose, 2% BSA, and 1% Triton X-100 in PBS) containing 1:500 rabbit anti-Lucifer Yellow antibody (Invitrogen A5750) and incubated gently for 3 days at 4 °C. Sections were washed three times for 5 min in blocking solution and incubated 2 h at room temperature with 1:400 biotinylated goat anti-rabbit antibody (Vector Laboratories BA-1000). Next, sections were washed three times for 5 min in PBS. A tertiary incubation was performed by incubating sections for 2 h at room temperature in streptavidin-conjugated Alexa 488 (Invitrogen S11223) diluted 1:1,000 in PBS. Finally, sections were washed three times in PBS, mounted on slides using Fluoro-Gel (EMS, 17985-10) and imaged by confocal microscopy. Spine density was calculated automatically using NeuronStudio (Mount Sinai School of Medicine) and manually curated by an observer using a three-dimensional analysis of the dendritic image stack. All spine counts began 30 μm away from the outer edge of the soma and extended for an additional 10–60 μm away from the starting point. The data from spine density passed the Lilliefors normality test and D'Agostino & Pearson omnibus normality test. Spine metrics relating to spine length, spine neck diameter and spine neck width were collected using ImageJ (NIH). All analyses of spine metrics were performed by observers that were blinded to the genotypes of the animals.

Electron microscopy. Mice were assigned a code previous to dissection, as to maintain a blinded genotype across all procedures, including dissection, sample processing, imaging and quantification. Mice were deeply anesthetized with an overdose of isofluorane and transcardially perfused with PBS (pH 7.4) followed by ice-cold 4% paraformaldehyde (PFA) in phosphate buffer (pH 7.4). The brains were removed, the striatum dissected and post-fixed overnight in PFA 4%, then transferred into a 4% glutaraldehyde solution and kept at 4 °C for 3 days. The samples were washed twice, 20 min each, in 7.5% sucrose, 0.1 M sodium cacodylate buffer, then post-fixed in 1% osmium tetroxide for 2 h with initial microwave treatment for 6 min. Next, the samples were washed twice in 0.11 M veronal acetate buffer for 20 min each. Following en-block staining in 1% uranyl acetate in distilled water for 1 h the samples were washed twice in 0.11 M veronal acetate buffer for 20 min each. Samples were dehydrated using serial dilutions of ethanol (70%, 95%, 2× 100%) for 20 min each, with initial microwave treatment of 2 min. Samples were then treated for 20 min twice with propaline oxide and impregnated with 50:50 propaline oxide:Epon resin overnight at 4 °C, with initial microwave treatment for 3 min. Next, the samples were impregnated with 100% Epon resin, three changes of 2 h each, with initial microwave treatment for 3 min each. Tissue samples were embedded in moulds and incubated for 48 h at 60 °C. Afterwards, semi-thin sections (0.5 μm) were cut on a Leica UltraCut S ultramicrotome and stained with Toluidine (0.8%) stain. From these, thin striatal sections (70 nm) were cut on an UltraCut S, mounted on 200 mesh Metaxaform Copper Rhodium grids and post-stained in 2% uranyl acetate in distilled water for 15 min and Sato's Lead citrate stain for 7 min. Grids were examined on a Philips (FEI) CM 12 transmission electron microscope. Images were acquired at ×40,000 magnification using an AMT 2Vue system, with an ORCA HR High resolution digital camera 7 megapixels, a Hamamatsu DCAM board for acquisition and AMT Image Capture Engine software version 600.335f. Images were saved as 7.5 megapixels 8 bit TIFF format files. PSD measurements were performed using ImageJ (NIH) by an observer that was blinded to the genotype of the samples.

Magnetic resonance image acquisition. Animals were assigned a blinding code, which was maintained during magnetic resonance (MR) data acquisition and analysis. MR mouse brain imaging was performed on a 7T Bruker Biospec 70/30 horizontal bore system (Billerica). Animals were lightly anesthetized under isofluorane with continuous monitoring and maintenance of physiological parameters throughout the imaging session (~60 min for each animal). Axial two-dimensional T2-weighted fast spin echo images (TURBO-RARE, TE/TR = 11/4,200 ms with 1 mm slice thick, matrix = 256 × 256 and FOV of 2.4 cm × 2.4 cm, five averages, 0.0 mm interslice gap) images were first obtained for screening purposes and supplemental anatomic information. For directed striatal and brain volumetric analysis, 64 contiguous 500-μm thick three-dimensional FSE proton density images (TURBO-RARE, TE/TR = 9/1,500 ms, matrix = 256 × 256 × 64 and FOV 2.2 cm × 2.2 cm × 2.2 cm, 25 min duration) were acquired.

MR volumetry measurements. Volumetric analysis of MR data sets was performed in OsiriX software, an open source image processing application

developed and maintained by Pixmeo. The left caudate-putamen was manually segmented in each animal by an investigator blinded to genotype. Each caudate-putamen was traced on contiguous axial slices from the three-dimensional volume acquisition with reference to a high-resolution age-matched mouse brain atlas (The Mouse Brain Atlas, The Mouse Brain Library at http://www.mbl.org). Selected areas were reviewed for consistency on coronal and sagittal representations, and cross-correlated with axial two-dimensional FSE images. Volumes were computed within OsiriX. Two separate striatal segmentations were obtained for each animal, with the average volume then taken. Intrarater reliability (kappa value) was = 0.97.

Volume normalization. Unilateral caudate-putamen volumes were normalized to brain volume measurements obtained from the same three-dimensional volume sets. Because of susceptibility distortions within the posterior fossa, and variable inclusion from animal to animal of posterior fossa contents at the caudal end of the 64 slice three-dimensional volume set, 'whole' brain volumes for normalization were obtained in each animal from traces beginning rostrally at the olfactory bulbs and ending caudally through the cerebral aqueduct at the roof of the fourth ventricle. As with striatal volumes, brain volumes were computed in OsiriX from the average of two segmentations. Intrarater reliability (kappa value) was >0.99. Statistical analysis was performed with unpaired two-tailed t-test.

39. Welch, J. M., Wang, D. & Feng, G. Differential mRNA expression and protein localization of the SAP90/PSD-95-associated proteins (SAPAPs) in the nervous system of the mouse. *J. Comp. Neurol.* **472,** 24–39 (2004).
40. Vorhees, C. V. & Williams, M. T. Morris water maze: procedures for assessing spatial and related forms of learning and memory. *Nature Protocols* **1,** 848–858 (2006).
41. Yin, H. H., Davis, M. I., Ronesi, J. A. & Lovinger, D. M. The role of protein synthesis in striatal long-term depression. *J. Neurosci.* **26,** 11811–11820 (2006).
42. Malenka, R. C. & Kocsis, J. D. Presynaptic actions of carbachol and adenosine on corticostriatal synaptic transmission studied *in vitro. J. Neurosci.* **8,** 3750–3756 (1988).
43. Myme, C. I., Sugino, K., Turrigiano, G. G. & Nelson, S. B. The NMDA-to-AMPA ratio at synapses onto layer 2/3 pyramidal neurons is conserved across prefrontal and visual cortices. *J. Neurophysiol.* **90,** 771–779 (2003).

ARTICLE

Local sleep in awake rats

Vladyslav V. Vyazovskiy[1], Umberto Olcese[1,2], Erin C. Hanlon[1], Yuval Nir[1], Chiara Cirelli[1] & Giulio Tononi[1]

In an awake state, neurons in the cerebral cortex fire irregularly and electroencephalogram (EEG) recordings display low-amplitude, high-frequency fluctuations. During sleep, neurons oscillate between 'on' periods, when they fire as in an awake brain, and 'off' periods, when they stop firing altogether and the EEG displays high-amplitude slow waves. However, what happens to neuronal firing after a long period of being awake is not known. Here we show that in freely behaving rats after a long period in an awake state, cortical neurons can go briefly 'offline' as in sleep, accompanied by slow waves in the local EEG. Neurons often go offline in one cortical area but not in another, and during these periods of 'local sleep', the incidence of which increases with the duration of the awake state, rats are active and display an 'awake' EEG. However, they are progressively impaired in a sugar pellet reaching task. Thus, although both the EEG and behaviour indicate wakefulness, local populations of neurons in the cortex may be falling asleep, with negative consequences for performance.

While animals are awake, the eyes are usually open; they move around and respond to their surroundings. During sleep, the eyes close, behaviour stops and animals fail to respond to stimuli. Studies of brain activity also show major differences between an awake state and non-rapid-eye-movement (NREM) sleep, which makes up ~80% of all sleep. While awake, neurons in the cerebral cortex fire irregularly, their membrane potential is tonically depolarized and an EEG shows low-voltage, high-frequency activity. During NREM sleep, neurons become bistable owing to a decrease in the level of neuromodulators: their membrane potential oscillates between a depolarized 'up' state similar to that seen in an awake state and a hyperpolarized 'down' state during which they cease firing altogether[1]. These slow oscillations occur in a range between 0.1 Hz and 6 Hz and they are detectable in the form of multi-unit activity ('on' and 'off' periods) as well as in EEG slow waves[2].

By staying awake too long, one becomes tired and many studies have demonstrated attention lapses, poor judgement and frequent mistakes in various cognitive tasks, even when the subject may not feel particularly sleepy[3,4]. Moreover, an EEG shows some trace of the sleep/wake history: the longer a subject has been awake, the higher the spectral power in the slow-wave range (0.5–4 Hz) of the EEG in subsequent sleep[5], corresponding to larger and more frequent slow waves and to more intense and synchronous neuronal activity[2]. Local variations in cortical activity while awake are associated with local changes during subsequent sleep and with a sleep-dependent increase in task performance[6-8]. These changes are reversed progressively in the course of sleep[5]. The awake EEG also shows changes that reflect the duration of previous awake states, with power increasing in the theta range (5–7 Hz)[9-11]. Likewise, neuroimaging studies show changes in blood flow and metabolism after sleep deprivation, with some brain regions undergoing decreases in activation and others, increases in activation[12]. However, any changes in underlying neuronal activity are poorly understood.

Neurons can go offline during a prolonged awake state

To investigate neuronal activity during a prolonged awake state, we implanted a group of adult rats ($n = 11$) with 16-channel microwire arrays in deep layers of the frontal motor cortex and recorded both the local field potentials (LFPs) and local multi-unit activity[2] across periods of spontaneous sleep and awake states (Supplementary Information). As expected, the awake LFP was characterized by low-amplitude fast waves and theta waves, accompanied by irregular,

tonic multi-unit activity. This was readily distinguishable from the LFP of NREM sleep, in which high-amplitude slow waves occurred concomitantly with synchronous 'on' and 'off' periods at the level of multi-unit activity (Fig. 1a).

We then kept rats awake by supplying novel objects for four hours starting at light onset[2]. As expected, by the end of sleep deprivation, the LFP showed an approximately 30% increase in spectral power in the slow/theta range (2–6 Hz; Fig. 1b, left panel, Supplementary Fig. 1a). However, close inspection of the recordings revealed an occasional change in firing patterns (Fig. 1c, left panel): towards the end of sleep deprivation (SD4), neuronal activity sporadically showed brief periods of silence involving all or most of the recorded neurons; this occurred less frequently at the beginning of sleep deprivation (SD1). These short off periods in populations of neurons were often associated with slow/theta waves in the LFP. An opposite dynamic was observed during 6 h of recovery sleep, when the LFP showed a progressive decline in slow-wave activity (Fig. 1b, right panel; Supplementary Fig. 1b). At the beginning of sleep (S1), large LFP slow waves were associated with synchronous on–off oscillations in multi-unit activity (Fig. 1c, right panel). At the end of recovery sleep (S6), large slow waves became infrequent and multi-unit activity became sparse and irregular. Thus, at the level of neuronal firing, wakefulness under high sleep pressure occasionally resembles late NREM sleep, whereas low-pressure sleep may occasionally resemble wakefulness.

Notably, the number of off periods in the awake state increased significantly from SD1 to SD4 (57.7% ± 16.5%, Fig. 1d, left panel), indicating that the tendency of neurons to enter a 'sleep-like' mode increases with sleep pressure. The number of high-amplitude LFP 2–6 Hz waves also increased significantly, by 23.3% ± 5.2% (Fig. 1d, right panel). The initial number of off periods or LFP 2–6 Hz waves during SD1 correlated negatively with their increase from SD1 to SD4 (off periods, $R = -0.53$, $P < 0.1$; LFP waves, $R = -0.87$, $P < 0.0001$), consistent with a saturating increase of sleep pressure[5]. Again, an opposite dynamic was apparent during recovery sleep: off periods and high-amplitude LFP slow waves during sleep decreased significantly from S1 to S6, by 36.9% ± 11.2% and 59.5% ± 9.0%, respectively (Fig. 1e). Of note, off periods, as defined here, and 2–6 Hz LFP waves were also observed during baseline spontaneous awake states in all rats but their frequency was lower than that observed during the first hour of sleep deprivation (off periods: 7.1 ± 4.1 per min,

[1]Department of Psychiatry, University of Wisconsin-Madison, Madison, 6001 Research Park Boulevard, Wisconsin 53719, USA. [2]PERCRO Laboratory, Scuola Superiore Sant'Anna, 56217, Via Martiri, 11, Pisa, Italy.
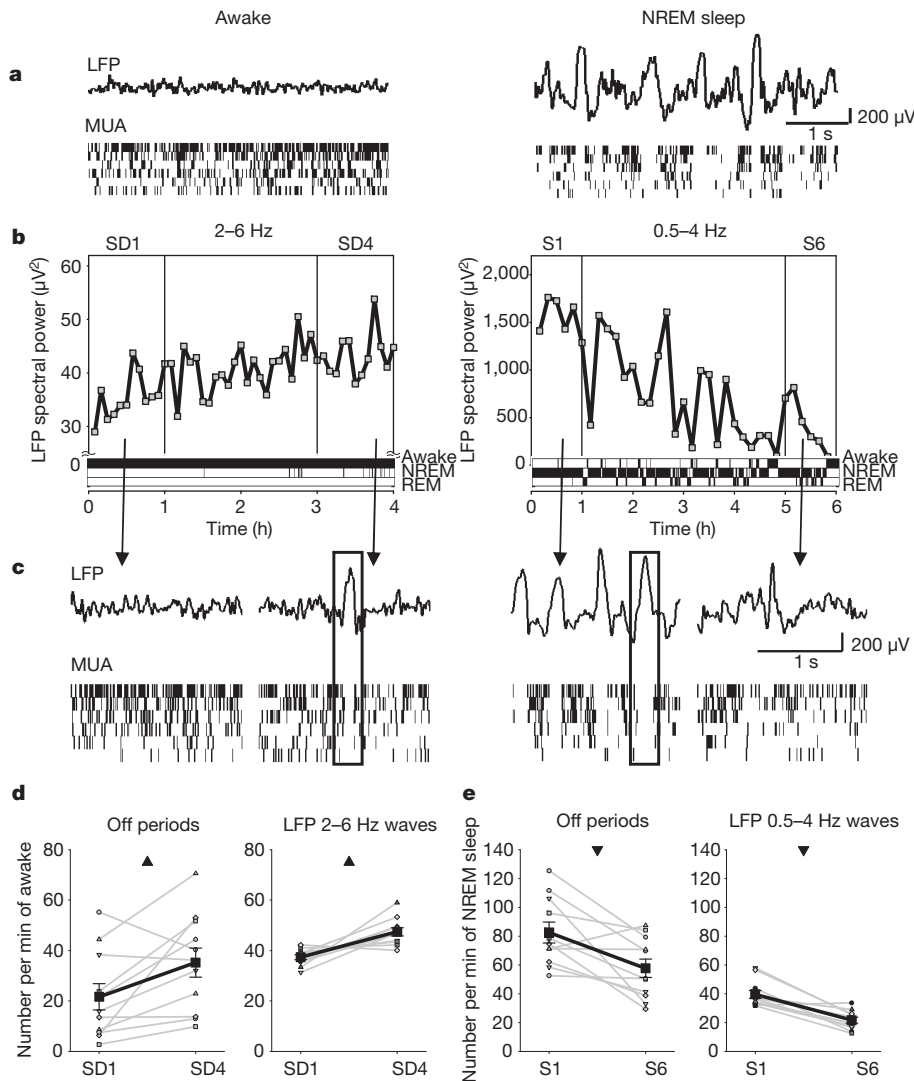
Figure 1 | Off periods during sleep and awake states. a, LFP records from the frontal cortex and raster plots of corresponding multi-unit activity (MUA; six putative neurons, each vertical line is a spike) in an awake state and in NREM sleep. **b**, Left panel: time course of LFP slow/theta power (2–6 Hz) for consecutive 5-min bins during 4 h of sleep deprivation in one rat. Right panel: time course of LFP slow-wave activity (0.5–4 Hz) for consecutive 5-min bins during 6 h of recovery after sleep deprivation in the same rat. Note the different y-axis scales in the two panels. Corresponding hypnograms are shown under each plot. **c**, LFP records and corresponding multi-unit activity raster plots in awake rats at the beginning (SD1) and end (SD4) of sleep deprivation (left panel) and during NREM sleep at the beginning (S1) and end (S6) of recovery (right panel). **d, e**, Changes in off periods and 2–6 Hz waves in awake rats (**d**) and in off periods and 0.5–4 Hz waves in NREM sleep (**e**). Black lines, mean ± s.e.m., n = 11 rats; grey lines, individual rats. Triangles depict significant differences (awake, off periods: $F_{(1,21)} = 7.03$, $P = 0.024$; 2–6 Hz LFP waves, $F_{(1,21)} = 18.61$, $P = 0.0015$; NREM sleep, off periods, $F_{(1,21)} = 10.40$, $P = 0.009$; 0.5–4 Hz LFP waves: $F_{(1,21)} = 34.83$, $P = 1.5069 \times 10^{-4}$, fixed-effects model analysis of variance (ANOVA)).

$P = 0.023$; 2–6 Hz waves, 25.05 ± 8.1 per min, $P = 0.06$). Thus, high sleep pressure is associated with an increased tendency of neurons to go offline in both awake and sleep states. Our data also indicate that off periods in multi-unit activity underlie the macroscopic changes in LFP low-frequency spectral power.

Asynchronous off periods in distant cortical regions

Sleep is usually considered a global behaviour and a global cortical and EEG state[13]. This raises the question of whether off periods during awake states can be detected simultaneously in distant cortical areas. In several animals ($n = 9$), we implanted an additional microwire array in the deep layers of the parietal cortex. We found that off periods in awake rats were also present in the parietal cortex (average duration: 79.02 ± 7.7 ms, incidence: 37.51 ± 6.16 per min) and that their occurrence increased from SD1 to SD4, similarly to the frontal off periods ($56.6\% \pm 19.5\%$, $n = 9$, $F_{(1,17)} = 6.23$, $P = 0.041$, fixed-effects model ANOVA). Moreover, during sleep deprivation, we found instances in which all recorded neurons in frontal and parietal areas underwent off periods near-simultaneously, consistent with this being a global phenomenon (Fig. 2a, left panel (global)). However, neurons recorded in one cortical area often showed an off period while neurons in the other area stayed on as they normally do in the awake state (Fig. 2a, left panel (local)) and most off periods were local, being observed only in one cortical region at a time (frontal, $76.9\% \pm 2.9\%$; parietal, $82.8\% \pm 3.1\%$; frontal versus parietal, $F_{(1,27)} = 4.6$, $P = 0.0981$). Notably, both global and local off periods

increased from SD1 to SD4 (Fig. 2b, left and middle panels) but the former increased more than the latter (Fig. 2b, right panel). Consistent with the multi-unit activity findings, most 2–6 Hz waves in awake rats occurred exclusively in the LFP from one of the two areas, whereas the remaining waves were seen near-simultaneously in both areas. Both patterns became more frequent from SD1 to SD4 but the relative proportion of global waves was greater in SD4 than in SD1 (Fig. 2d), indicating that as sleep pressure builds up, neuronal activity in an awake state becomes more synchronized, just as it does in sleep.

Having established that distant brain areas can enter off periods independently in behaviourally awake animals during sleep deprivation, we asked next if nearby neurons (separated by ~2 mm) can also do so. We found that, even among units recorded with the same microelectrode array, a substantial fraction of neurons could stop firing together for up to hundreds of milliseconds while the remaining neurons maintained their spiking activity at virtually unaltered or even elevated rates (Supplementary Fig. 2a). On average, while a subset of neurons ceased firing abruptly, the firing of the remaining neurons increased transiently with a delay of ~20 ms and then slowed down by ~15% (Supplementary Fig. 2b). Controls conducted by shuffling units between subsets of neurons indicated that these hyper-local off periods were unlikely to be an artefact of different firing rates of cortical neurons (Supplementary Information). Hyper-local off periods increased by almost 40% from SD1 to SD4, indicating that they too are related to sleep pressure (Supplementary Fig. 2c).

Having found evidence for local off periods during the awake state, we then asked if off periods during sleep could also be local. Previous
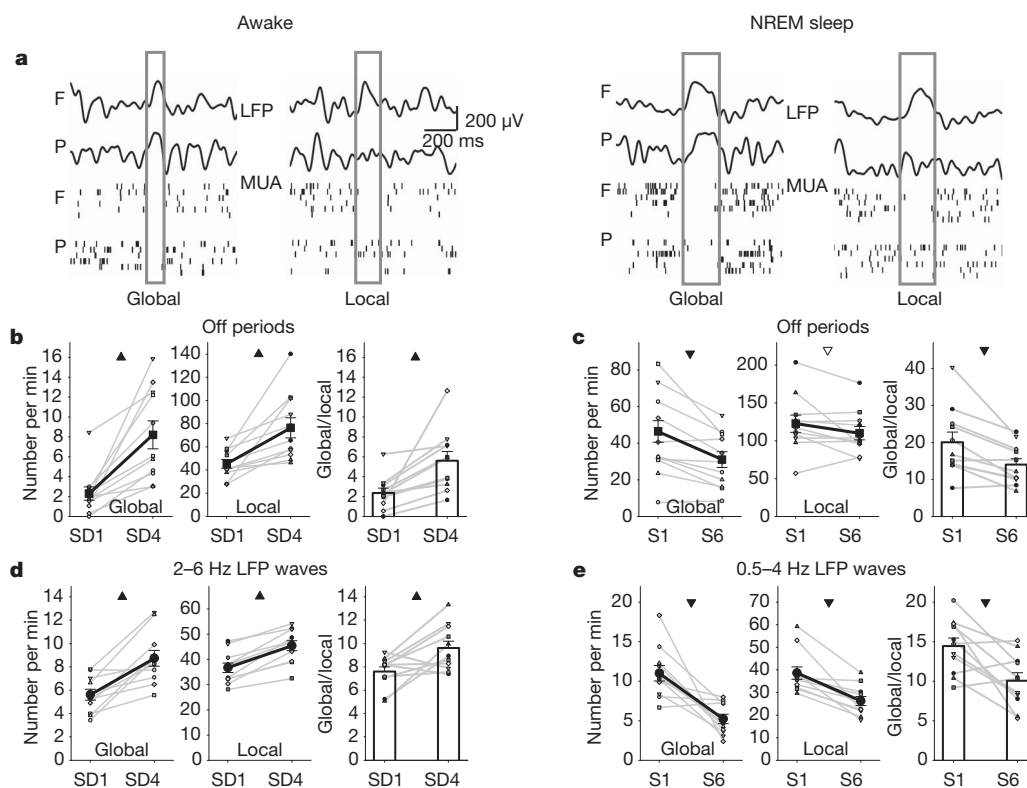
Figure 2 | Local off periods in awake rats. a, Left panels: Awake LFP records in frontal (F) and parietal (P) cortices, depicting global or local frontal 2–6 Hz waves (boxed) and raster plots of corresponding multi-unit activity. Right panels: LFP records in NREM sleep depicting global or local slow waves (boxed) and raster plots of corresponding multi-unit activity. **b–e**, Left and middle panels: changes in global and local off periods/LFP waves (F+P/2) during sleep deprivation and recovery sleep. Black lines, mean ± s.e.m., $n = 7$ rats, 1–3 experiments per rat; grey lines, individual rats. Right panels: number of global off periods/LFP waves as a percentage of local off periods/LFP waves (F+P). Triangles depict differences at a significant (filled) or tendency (open) level. **b**, Global off periods in awake rats: $F_{(1,21)} = 94.95$, $P = 0.0104$; local off periods in awake rats: $F_{(1,21)} = 20.08$, $P = 0.0464$. Right panel: global off periods as a percentage of local during SD1 and SD4 ($F_{(1,21)} = 67.05$, $P = 0.0146$, fixed-effects model ANOVA). **c**, Global off periods in NREM sleep: $F_{(1,21)} = 60.72$, $P = 0.0161$; local off periods in NREM sleep: $F_{(1,21)} = 11.56$, $P = 0.0767$. Right panel: global off periods as a percentage of local during S1 and S6 ($F_{(1,21)} = 99.17$, $P = 0.0099$, fixed-effects model ANOVA). Note the different y-axis scales in **b** and **c**. **d**, Global waves in awake rats: $F_{(1,21)} = 34.08$, $P = 0.0281$; local waves in awake rats: $F_{(1,21)} = 28.54$, $P = 0.0333$. Right panel: global waves as a percentage of local during SD1 and SD4 ($F_{(1,21)} = 52.53$, $P = 0.0185$, fixed-effects model ANOVA). **e**, Global waves in NREM sleep: $F_{(1,21)} = 254.42$, $P = 0.0039$; local waves in NREM sleep: $F_{(1,21)} = 529.31$, $P = 0.0019$. Right panel: global waves as a percentage of local during S1 and S6 ($F_{(1,21)} = 37.38$, $P = 0.0253$, fixed-effects model ANOVA). Note the different y-axis scales in **d** and **e**.

evidence has shown that sleep can be regulated locally[14], as demonstrated by a local increase in slow-wave activity after manipulations that affect neuronal plasticity during the awake state[8]. Moreover, high-density EEG studies in humans combined with source localization[15], as well as modelling studies[16], have indicated that slow waves with multiple peaks during sleep may result from the summation or interference of separate slow waves originating at different locations. Finally, recent depth recordings in humans have provided evidence that sleep slow waves and off periods can be local[17]. As shown in Fig. 2a, right panel, we found that in rats, off periods during NREM sleep occurred not only synchronously at frontal and parietal areas but also locally, in which case they were associated with local slow waves in the LFP (Fig. 2a, right panel (local)). The incidence of both global and local off periods in NREM sleep decreased significantly from S1 to S6 (Fig. 2c, left and middle panels), accompanied by a relative reduction of global slow waves (Fig. 2c, right panel). Thus, just as 2–6 Hz waves in awake rats became more global from SD1 to SD4 (Fig. 2b, right panel), sleep slow waves became more local from S1 to S6 (Fig. 2c, right panel), indicating that populations of neurons are more easily recruited into synchronous slow oscillations when sleep pressure is high than when it has dissipated[2,18].

Local off periods in an awake state lead to behavioural deficits

It is common experience that tiredness after prolonged sleep deprivation can be manifested as 'microsleeps': brief episodes of 3–15 s during which a person appears suddenly asleep (eyes closed or closing), may not respond to stimuli and shows sleep-like EEG activity[19]. Clearly, such microsleeps can be dangerous during tasks requiring alertness and the detection of sleep-like behaviour or EEG changes is being pursued to reduce risks[20]. However, careful observation of our rats, which were exposed to a relatively short period of sleep deprivation, did not reveal any indication of sleep: their eyes were open, they responded to stimuli and their EEG was unambiguously an awake EEG (Fig. 1a, Supplementary Information and Supplementary Figs 3 and 4). Moreover, a retrospective analysis of video recordings showed no behavioural signs of sleep specifically during multi-unit activity off periods.

Given that the off periods we detected have no overt manifestations, are brief and are often local, this raises the question of whether they have any impact on performance. To investigate the potential consequences of neuronal 'tiredness', rats were trained on a sugar pellet reaching task[21] for 2 h between SD1 and SD4 (Supplementary Information). Learning the reaching task engages a circumscribed cortical area in the motor cortex and leads to local plasticity while awake[22] and to increased slow-wave activity during subsequent sleep[21,23]. To investigate directly whether an increased incidence of neuronal off periods leads to impaired performance, we conducted simultaneous video and multi-unit activity recordings with high temporal resolution in a subset of animals ($n = 8$) during the reaching attempts. Although the number of off periods decreased steadily in both the frontal and parietal cortices towards each reaching attempt, possibly reflecting the increased global arousal necessary to perform the reach, we found that the occurrence of an off period within several
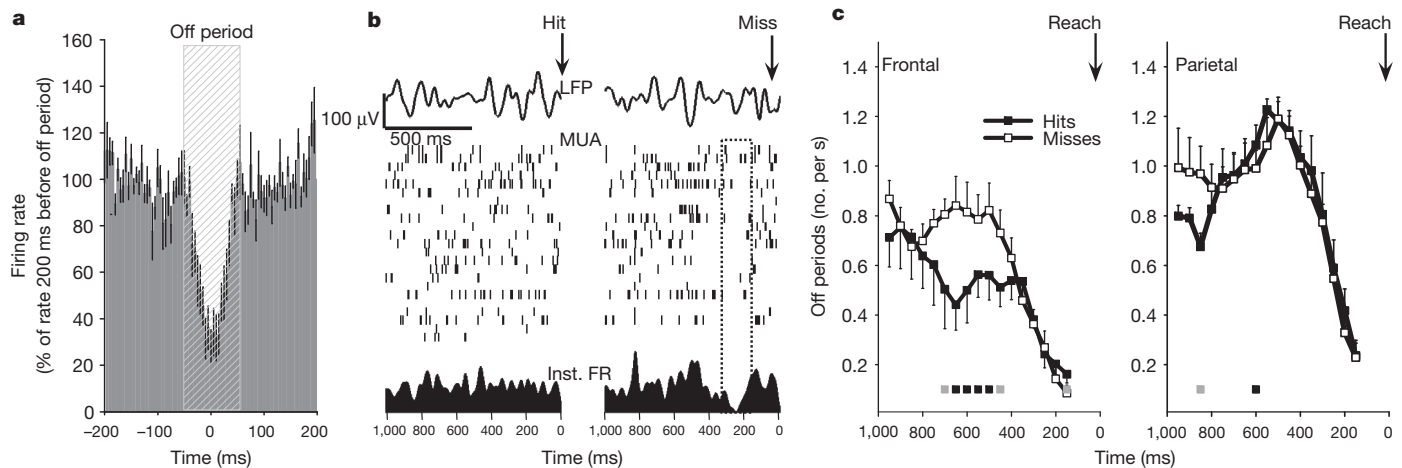
Figure 3 | Off periods in awake rats affect performance. a, Average neuronal activity in the frontal cortex triggered by off periods (mean ± s.e.m., $n = 6$ rats; shown as a percentage of mean firing rate in the last 200 ms before the off period). **b**, Top: individual examples of frontal LFP records immediately preceding a successful or unsuccessful reaching attempt ('Hit' or 'Miss', arrows); middle, raster plots of corresponding multi-unit activity; bottom: instantaneous firing rates (Inst. FR) of the entire population (20 putative neurons). Note a generalized suppression of firing before a miss (boxed area). **c**, Numbers of off periods before hits or misses (frontal, $n = 6$ rats; parietal, $n = 5$ rats). Average values are plotted for consecutive overlapping 300-ms windows with 50-ms shifts against the midpoint of the corresponding window (for example, the value at 500 ms depicts the number of off periods occurring between 350 and 650 ms). Squares show significant differences between hits and misses (grey, $P < 0.1$; black, $P < 0.05$; rANOVA).

hundred milliseconds before the reaching attempt was often associated with failure to successfully grasp a sugar pellet (Fig. 3a, b). Specifically, off periods occurred more frequently ~300–800 ms before an unsuccessful reaching attempt as compared to successful trials (Fig. 3c, left panel) and the probability of a successful reach decreased by 37.5% if there was at least one off period before the reach (off+: $26.1 ± 6.3$, off−: $41.8 ± 4.1$; $F_{(1,11)} = 15.6$, $P = 0.01$). Importantly, this effect was observed in the frontal but not in the parietal cortex (Fig. 3c, right panel). We also found that the overall number of misses increased significantly across the training periods ($P < 0.05$) and that behaviour became progressively more unstable. At the beginning of the task, hits and misses alternated regularly, but as time progressed, long clusters of misses became more frequent and had increasingly variable duration (duration, $F_{(2,32)} = 4.69$, $P = 0.021$; variance, $F_{(2,32)} = 4.31$, $P = 0.028$, repeated measures ANOVA). These results indicate that neuronal off periods and corresponding increases in low-frequency LFP power may be associated with decreased behavioural performance, as is typical of sleep-deprived individuals.

Discussion

These findings show that, in animals kept awake beyond their normal sleep time, populations of neurons in different cortical areas can suddenly go 'offline' in a way that resembles the off periods of NREM sleep. The main differences are that during sleep, virtually all cortical neurons show on–off oscillations in the slow-wave frequency range, the EEG displays typical sleep slow waves and spindles and the animal is behaviourally immobile and unresponsive, with eyes closed. During a prolonged awake state, however, only subsets of neurons enter off periods, usually for shorter durations; the EEG is typical of an awake state and the animal appears behaviourally awake with eyes open and is responsive to stimuli. Furthermore, the number of off periods increases with the duration of wakefulness, indicating that the likelihood of subsets of neurons going offline in an otherwise-awake cortex increases with sleep pressure. As shown here, the progressive changes observed during sleep deprivation are the mirror image of changes during recovery sleep: neuronal firing rates during on periods, the number and duration of off periods, the number of neurons participating synchronously in off periods and the low-frequency content of the EEG all increase during an awake state just as they decrease during sleep. This supports the concept of homeostatic regulation of the need for sleep[5].

Perhaps the most striking result of this study is that in the sleep-deprived brain, subsets of neurons may enter an off period in one

cortical area but not in another and that even within the same cortical area, some neurons may be off while others remain on. On the basis of this evidence, the wake behaviour of a sleep-deprived subject might be characterized as a covert form of 'dormiveglia' or 'sleep/wake'[24]. Moreover, as shown here using the sugar pellet reaching task, the increasing occurrence of local off periods during a prolonged period of being awake was associated with worsening performance in the task. Paradigms should be developed to associate the occurrence of off periods in specific subsets of neurons more precisely with specific performance failures but these initial findings raise the intriguing possibility that 'local sleep' in an awake brain may be responsible for cognitive impairments due to sleep deprivation or restriction[3,4]. It is especially relevant that cognitive impairments, including defective judgment and irritability[25], may occur despite an outward impression of wakefulness, the lack of subjective insight[4] and an awake EEG. Sporadic, local neuronal off periods in sleep-deprived subjects may be analogous to the sporadic, local, hyper-synchronous discharges seen in partial epilepsy. Such local events can be detected with careful EEG recordings as interictal spikes and may cause momentary lapses (absence) without overt behavioural signs[26].

We can only speculate about the mechanisms underlying the local awake off periods. A spontaneous slow oscillation of membrane potentials can occur in the mouse barrel cortex during quiet wakefulness[27] and can affect the amplitude of evoked responses[14], although it is not clear whether such 'down' states occur during active behaviour, are local, affect performance and most importantly, reflect increasing sleep pressure. Although we do not know whether the awake off periods we observed in freely moving rats are associated with neuronal hyper-polarization, their overall similarity to sleep off periods, and the finding that they become more frequent with increasing sleep pressure, indicates that they may be an expression of increasing bistability in neurons[28]. Thus, in addition to the global state of instability that is a hallmark of sleep deprivation[29], there can also be a local instability, at least in the cerebral cortex. Bistability between on and off periods[30] could be triggered by decreasing levels of arousal-promoting neuromodulators[28], especially because cholinergic and noradrenergic neurons, for instance, do not always discharge in tight synchrony[31,32] and presynaptic release can be modulated locally[33,34].

Local sleep in awake rats may be either an adaptive or a maladaptive response. In some cetaceans and birds, one hemisphere can remain awake while the other is in slow wave sleep, an adaptive response

that permits them to continue swimming, flying or monitoring the environment[35]. The ability to control behaviour actively with some neural circuits while others may be idling[36] could be evolutionarily advantageous. However, dissociated behavioural states, such as sleep-walking, REM sleep behaviour disorder and other parasomnias, are clearly maladaptive[37,38]. Because local awake off periods are associated with locally increased excitability after intensive training and with failures in performance, it is likely that they represent a form of neuronal tiredness due to use-dependent factors, such as synaptic overload[39]. A question for the future is whether local off periods in the awake state may also serve functional roles, from energy saving[39,40] to the initiation of a local restorative process.

METHODS SUMMARY

In male WKY rats, LFPs and multi-unit activity were recorded from deep layers of the frontal ($n = 11$ rats) and/or the parietal ($n = 9$) cortex with 16-channel (2×8) polyimide-insulated tungsten micro-wire arrays. Rats were housed individually in transparent Plexiglas cages (light:dark 12:12 h, light on at 10:00; 23 ± 1 °C; food and water ad libitum and replaced daily at 10:00 except for the sugar pellet reaching task; see Supplementary Information). Animal protocols followed the National Institutes of Health guide for the care and use of laboratory animals and were in accordance with institutional guidelines.

Data acquisition and online spike sorting were performed with the Multichannel Neurophysiology Recording and Stimulation System (Tucker-Davis Technologies Inc). Multi-unit activity was collected continuously (25 kHz, 300–5,000 Hz), concomitantly with the LFPs from the same electrodes and epidural EEGs (both 256 Hz, 0.1–100 Hz). Amplitude thresholds for online spike detection were set manually and only allowed crossings of spikes below $-25\,\mu V$. LFP power spectra were computed by a Fast Fourier Transform routine for 4-s epochs (Hanning window, 0.25 Hz resolution). Sleep stages were scored offline by visual inspection of 4-s epochs, in which the EEG, LFP, electromyogram (EMG) and spike activity were displayed. Spike sorting was performed by principal component analysis (PCA) followed by split and merge expectation maximization (SMEM) clustering algorithm. Population off periods in the awake state and during NREM sleep were defined as periods with suppressed or absent neuronal activity. Recordings were performed continuously for two to three weeks. In each animal, two to four experiments with 4 h of sleep deprivation were performed (at least 5 days apart), one of which was combined with the sugar pellet reaching task. For details about the analysis of firing rates and neuronal population off periods, see Supplementary Information.

1. Steriade, M., Timofeev, I. & Grenier, F. Natural waking and sleep states: a view from inside neocortical neurons. *J. Neurophysiol.* **85,** 1969–1985 (2001).
2. Vyazovskiy, V. V. et al. Cortical firing and sleep homeostasis. *Neuron* **63,** 865–878 (2009).
3. Dijk, D. J., Duffy, J. F. & Czeisler, C. A. Circadian and sleep/wake dependent aspects of subjective alertness and cognitive performance. *J. Sleep Res.* **1,** 112–117 (1992).
4. Van Dongen, H. P., Maislin, G., Mullington, J. M. & Dinges, D. F. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep* **26,** 117–126 (2003).
5. Borbély, A. A. & Achermann, P. Sleep homeostasis and models of sleep regulation. In *Principles and Practice of Sleep Medicine* (eds Kryger M. H., Roth, T., & Dement, W. C.) 405–417 (W. B. Saunders, 2005).
6. Dang-Vu, T. T. et al. Functional neuroimaging insights into the physiology of human sleep. *Sleep* **33,** 1589–1603 (2010).
7. Diekelmann, S. & Born, J. The memory function of sleep. *Nature Rev. Neurosci.* **11,** 114–126 (2010).
8. Huber, R., Ghilardi, M. F., Massimini, M. & Tononi, G. Local sleep and learning. *Nature* **430,** 78–81 (2004).
9. Finelli, L. A., Baumann, H., Borbely, A. A. & Achermann, P. Dual electroencephalogram markers of human sleep homeostasis: correlation between theta activity in waking and slow-wave activity in sleep. *Neuroscience* **101,** 523–529 (2000).
10. Vyazovskiy, V. V. & Tobler, I. Theta activity in the waking EEG is a marker of sleep propensity in the rat. *Brain Res.* **1050,** 64–71 (2005).
11. Leemburg, S. et al. Sleep homeostasis in the rat is preserved during chronic sleep restriction. *Proc. Natl Acad. Sci. USA* **107,** 15939–15944 (2010).
12. Maquet, P. Functional neuroimaging of normal human sleep by positron emission tomography. *J. Sleep Res.* **9,** 207–231 (2000).
13. Sejnowski, T. J. & Destexhe, A. Why do we sleep? *Brain Res.* **886,** 208–223 (2000).
14. Krueger, J. M. et al. Sleep as a fundamental property of neuronal assemblies. *Nature Rev. Neurosci.* **9,** 910–919 (2008).
15. Riedner, B. A. et al. Sleep homeostasis and cortical synchronization: III. A high-density EEG study of sleep slow waves in humans. *Sleep* **30,** 1643–1657 (2007).
16. Esser, S. K., Hill, S. L. & Tononi, G. Sleep homeostasis and cortical synchronization: I. Modeling the effects of synaptic strength on sleep slow waves. *Sleep* **30,** 1617–1630 (2007).
17. Nir, Y. et al. Regional slow waves and spindles in human sleep. *Neuron* doi: 10.1016/j.neuron.2011.02.043.
18. Vyazovskiy, V. V., Faraguna, U., Cirelli, C. & Tononi, G. Triggering slow waves during NREM sleep in the rat by intracortical electrical stimulation: effects of sleep/wake history and background activity. *J. Neurophysiol.* **101,** 1921–1931 (2009).
19. Tirunahari, V. L., Zaidi, S. A., Sharma, R., Skurnick, J. & Ashtyani, H. Microsleep and sleepiness: a comparison of multiple sleep latency test and scoring of microsleep as a diagnostic test for excessive daytime sleepiness. *Sleep Med.* **4,** 63–67 (2003).
20. Blaivas, A. J., Patel, R., Hom, D., Antigua, K. & Ashtyani, H. Quantifying microsleep to help assess subjective sleepiness. *Sleep Med.* **8,** 156–159 (2007).
21. Hanlon, E. C., Faraguna, U., Vyazovskiy, V. V., Tononi, G. & Cirelli, C. Effects of skilled training on sleep slow wave activity and cortical gene expression in the rat. *Sleep* **32,** 719–729 (2009).
22. Rioult-Pedotti, M. S., Friedman, D. & Donoghue, J. P. Learning-induced LTP in neocortex. *Science* **290,** 533–536 (2000).
23. Vyazovskiy, V. V. & Tobler, I. Handedness leads to interhemispheric EEG asymmetry during sleep in the rat. *J. Neurophysiol.* **99,** 969–975 (2008).
24. Moruzzi, G. (ed.) The Functional Significance of Sleep with Particular Regard to the Brain Mechanisms Underlying Consciousness. (Springer, 1966).
25. Killgore, W. D. Effects of sleep deprivation on cognition. *Prog. Brain Res.* **185,** 105–129 (2010).
26. Ferrillo, F. et al. Sleep-EEG modulation of interictal epileptiform discharges in adult partial epilepsy: a spectral analysis study. *Clin. Neurophysiol.* **111,** 916–923 (2000).
27. Poulet, J. F. & Petersen, C. C. Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice. *Nature* **454,** 881–885 (2008).
28. Hill, S. & Tononi, G. Modeling sleep and wakefulness in the thalamocortical system. *J. Neurophysiol.* **93,** 1671–1698 (2005).
29. Doran, S. M., Van Dongen, H. P. & Dinges, D. F. Sustained attention performance during sleep deprivation: evidence of state instability. *Arch. Ital. Biol.* **139,** 253–267 (2001).
30. Chauvette, S., Volgushev, M. & Timofeev, I. Origin of active states in local neocortical networks during slow sleep oscillation. *Cereb. Cortex* **20,** 2660–2674 (2010).
31. Wu, M. F. et al. Locus coeruleus neurons: cessation of activity during cataplexy. *Neuroscience* **91,** 1389–1399 (1999).
32. Nunez, A. Unit activity of rat basal forebrain neurons: relationship to cortical activity. *Neuroscience* **72,** 757–766 (1996).
33. Laplante, F., Morin, Y., Quirion, R. & Vaucher, E. Acetylcholine release is elicited in the visual cortex, but not in the prefrontal cortex, by patterned visual stimulation: a dual in vivo microdialysis study with functional correlates in the rat brain. *Neuroscience* **132,** 501–510 (2005).
34. Marrocco, R. T., Lane, R. F., McClurkin, J. W., Blaha, C. D. & Alkire, M. F. Release of cortical catecholamines by visual stimulation requires activity in thalamocortical afferents of monkey and cat. *J. Neurosci.* **7,** 2756–2767 (1987).
35. Cirelli, C. & Tononi, G. Is sleep essential? *PLoS Biol.* **6,** e216 (2008).
36. Pigarev, I. N., Nothdurft, H. C. & Kastner, S. Evidence for asynchronous development of sleep in cortical areas. *Neuroreport* **8,** 2557–2560 (1997).
37. Terzaghi, M. et al. Evidence of dissociated arousal states during NREM parasomnia from an intracerebral neurophysiological study. *Sleep* **32,** 409–412 (2009).
38. Mahowald, M. W. & Schenck, C. H. Insights from studying human sleep disorders. *Nature* **437,** 1279–1285 (2005).
39. Tononi, G. & Cirelli, C. Sleep function and synaptic homeostasis. *Sleep Med. Rev.* **10,** 49–62 (2006).
40. Attwell, D. & Laughlin, S. B. An energy budget for signaling in the grey matter of the brain. *J. Cereb. Blood Flow Metab.* **21,** 1133–1145 (2001).

ARTICLE

Structure and mechanism of the chromatin remodelling factor ISW1a

Kazuhiro Yamada[1]*, Timothy D. Frouws[1]*, Brigitte Angst[1]*, Daniel J. Fitzgerald[1]†, Carl DeLuca[1]†, Kyoko Schimmele[1], David F. Sargent[1] & Timothy J. Richmond[1]

Site-specific recognition of DNA in eukaryotic organisms depends on the arrangement of nucleosomes in chromatin. In the yeast *Saccharomyces cerevisiae*, ISW1a and related chromatin remodelling factors are implicated in establishing the nucleosome repeat during replication and altering nucleosome position to affect gene activity. Here we have solved the crystal structures of *S. cerevisiae* ISW1a lacking its ATPase domain both alone and with DNA bound at resolutions of 3.25 Å and 3.60 Å, respectively, and we have visualized two different nucleosome-containing remodelling complexes using cryo-electron microscopy. The composite X-ray and electron microscopy structures combined with site-directed photocrosslinking analyses of these complexes suggest that ISW1a uses a dinucleosome substrate for chromatin remodelling. Results from a remodelling assay corroborate the dinucleosome model. We show how a chromatin remodelling factor could set the spacing between two adjacent nucleosomes acting as a 'protein ruler'.

DNA in eukaryotic cells is packaged in nucleosomes to comprise chromatin, serving generally to protect DNA integrity and to repress gene transcription. Chromatin structure depends on regular arrays of nucleosomes for the formation of a compact chromatin fibre, as demonstrated *in vitro*[1–3]. Conversely, nuclear processes involving DNA depend on nucleosome mobility and the alteration of chromatin structure. Recent studies of nucleosome occupancy *in situ* show that nucleosome positional regularity is characteristic of chromatin, with distinct patterns appearing in the vicinity of gene promoter elements[1–6]. Over the past decade, four families of ATP-dependent chromatin remodelling factors (INO80, ISWI, Mi2/CHD and SWI/SNF) have been discovered that actively space, translocate or disassemble nucleosomes[7]. The imitation switch family (ISWI) is represented by the well-characterized yeast factors ISW1a, ISW1b and ISW2, human ACF/CHRAC, and *Drosophila* NURF[8–12]. ISWI members are important for regularization of nucleosome spacing and modulation of gene transcription[13–16]. Nevertheless, little structural information exists to support potential mechanisms of chromatin remodelling.

We have determined the crystal structure of ISW1a(ΔATPase)—that is, ISW1a lacking its ATPase domain—alone and with two copies of a 48-base-pair (bp) DNA bound at resolutions of 3.25 Å and 3.6 Å, respectively. ISW1a comprises the subunits Isw1 and Ioc3. Isw1 contains an SF2 ATPase domain (655 amino acids) connected by a linker (133 amino acids) to the subdomains HAND, SANT and SLIDE (hereafter referred to as HSS). The ATPase domain shares ~30% identity with *Sulfolobus solfataricus* Swi2/Snf2 and zebrafish Rad54 ATPase core domains for which there are high-resolution structures[17,18]. These studies on ATPase domains address questions primarily about translocation along naked DNA. We removed the ATPase domain and linker region of Isw1 to facilitate crystallization of the HSS–Ioc3 complex. The X-ray structure of the *Drosophila melanogaster* HSS domain was determined previously at 2.5 Å resolution, and a model was proposed for DNA binding based on homology with the cMyb-SANT domain[19]. Isw1 and *Drosophila* HSS are 37% homologous.

In addition to the ISW1a(ΔATPase) crystal structures, we report here cryo-electron microscopy (cryo-EM) structures and site-directed

photochemical mapping for two different ISW1a(ΔATPase)–nucleosome complexes. Our results suggest that ISW1a can bind two nucleosomes simultaneously. The ensuing HSS–Ioc3–dinucleosome model shows how ISW1a could set the spacing between two adjacent nucleosomes using HSS–Ioc3 as a protein ruler. Site-directed mapping of these nucleosomes in an ATP-dependent remodelling assay demonstrates that one nucleosome is static and the other mobile.

X-ray structure of ISW1a(ΔATPase)

The structure of ISW1a(ΔATPase), or equivalently HSS–Ioc3, is 'L'-shaped with dimensions of approximately $130 \times 80 \times 60$ Å (Fig. 1a). Ioc3 is a novel structure (based on DaliLite[20]) containing 16 α-helices, 12 of which surround a short four-stranded β-sheet in a globular core domain (Ioc3 core). As designated here, the helical-linker-DNA-binding domain (HLB) comprising α-helices α8–α11 protrudes from the core, and together account for the width of HSS–Ioc3 (Fig. 1b and Supplementary Figs 1 and 2). An HSS-binding loop (HSSB, 45 amino acids) also extends from the core between α5 and α6. The elongated HSS is attached only by SLIDE to the Ioc3 core, and together they account for the length of HSS–Ioc3. HSS in complex with Ioc3 has secondary structural elements identical to the previous structures alone from fly and frog[19,21], suggesting that the HSS has a rigid structure throughout the ISWI family.

The HSS–Ioc3 interface also appears to be rigid. SLIDE helices α9 and α10 are cradled in a large pocket formed primarily by Ioc3 α1, α6, α7 and α13, and HSSB projects across SLIDE α9 to reach HSS α7 and SANT (Supplementary Fig. 3). The solvent-accessible surface area buried between HSS and Ioc3 is substantial (5,303 Å², probe radius 1.4 Å), and is contributed by two hydrophobic regions comprising 26 side chains from the Ioc3 core and SLIDE, and a more distributed set of 11 side chains from HSSB, SANT and the SANT–SLIDE connector. A network of hydrogen bonds that includes 20 side chains, many charged, divides the Ioc3/SANT hydrophobic cluster and extends beyond it. The second hydrophobic region is also augmented by
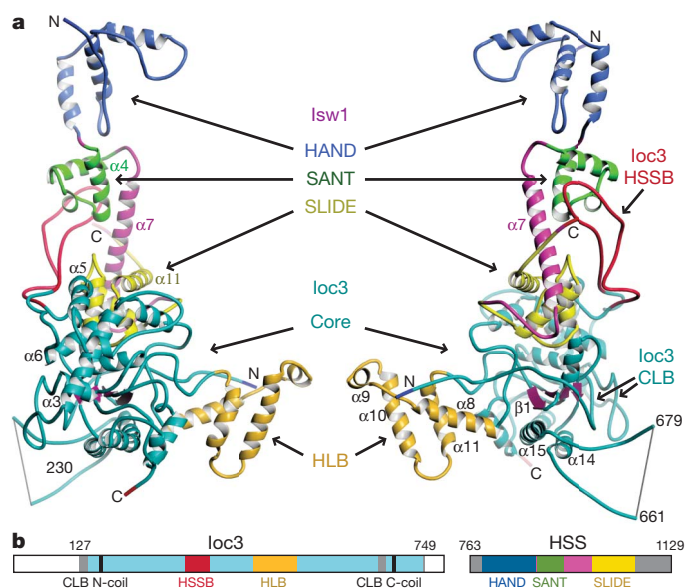
Figure 1 | ISW1a(ΔATPase) X-ray structure. a, Two views from opposite sides with regions of the subunits Ioc3 and Isw1(ΔATPase) indicated.
b, Sequence schematics of the protein constructs used for the crystal structure. Regions in colour were used for the structure (grey denotes unobserved). For Ioc3, the core domain (cyan), the HLB domain (gold), the two regions of the CLB motif (black) and the HSSB loop (red) are indicated. For Isw1(ΔATPase) (magenta), subdomains of the HAND–SANT–SLIDE domain (HSS) are marked. The ATPase domain and linker to HAND are the N-terminal 762 amino acids of Isw1 (not shown).

hydrogen bonds involving ten side chains. Main-chain amide and carbonyl groups are also part of these hydrogen bonding schemes.

The yeast protein Esc8, implicated in gene silencing through genetic and physical interaction with Sir2 histone deacetylase, shares significant sequence homology with Ioc3 (ref. 22). Esc8 was suggested as a partner for Isw1 based on homology and a survey of protein interactions in yeast[23]. The conservation between the putative Isw1–Esc8 and observed Isw1–Ioc3 interfaces strongly suggests that Esc8 is a chromatin-remodelling-factor subunit (Supplementary Fig. 4).

X-ray structure with DNA

HSS–Ioc3 displays two pronounced regions of positive electrostatic surface charge important for DNA binding (Supplementary Fig. 5). The structure of ISW1a(ΔATPase)–DNA reveals two 48-bp DNA duplexes bound separately to these regions (Fig. 2a). Ioc3 binds the first copy (E-DNA: related to external-linker DNA, see dinucleosome model) at the sequence $5'-T_{-1}A_1G_2G_3C_4T_5-3'$ (opposite strand: $3'-A'_{-1}T'_1C'_2C'_3G'_4A'_5-5'$) containing the DNA palindromic centre at base-pair step $T_{-1}A_1$ (Fig. 2b) coincident with a crystallographic two-fold axis. As a consequence, these central two base pairs are bound by two copies of Ioc3 (not shown). The helix axis of the 10-bp segment between the copies is bent by ~55°, and the trajectory of the remaining DNA may be affected by further crystal-packing interactions. Crystals containing iodoT at either 5 or 9 bp from the DNA centre show anomalous-difference electron density at the expected two-fold-related positions, confirming our DNA sequence assignment at this site (not shown). Two Ioc3 chain segments termed N-coil (amino acids 160–167) and C-coil (amino acids 691–696) are in close proximity with the E-DNA and together comprise the coil-linker-DNA-binding motif (CLB). CLB interacts primarily with DNA phosphate groups: C-coil T698-Oγ with T_{-1}, C-coil K692-Nε and amide-N with C'_3 and G'_4, respectively, N-coil Q161-Nε with A_1, and N-coil R167 guanidinium group with A'_5. Importantly, the methyl groups of C-coil A694 and T_{-1} base make a possible cognate interaction. Two further side chains, K695 and D693, project into the major groove and are aligned with base pair $G_4C'_4$, but are 0.1–0.6 Å too distant from base edges to make
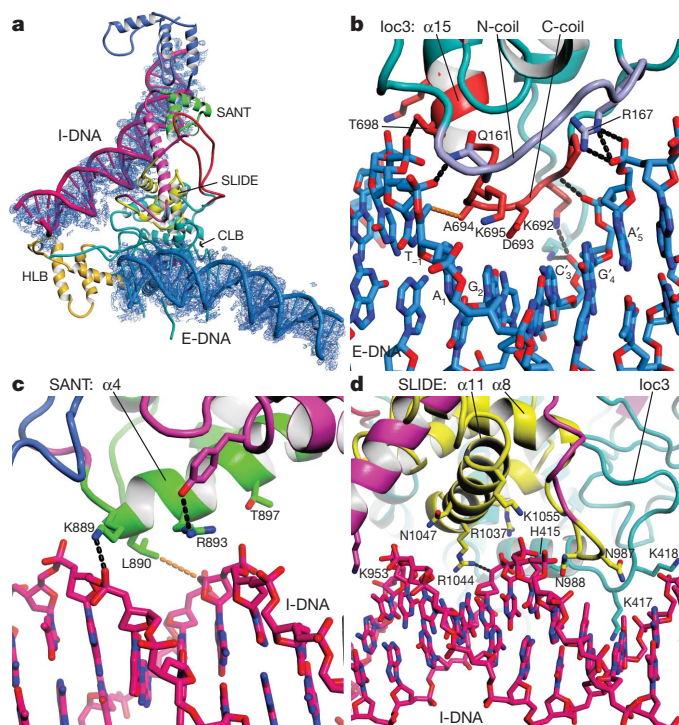

Figure 2 | ISW1a(ΔATPase)–DNA X-ray structure. a, Complex including the E-DNA (blue) bound to CLB, and the I-DNA (red) bound to SANT (green) and SLIDE (yellow) and terminating at HLB (gold). The 3.6 Å resolution, $2F_o − F_c$ electron density (blue) contoured at $1.0σ$ is shown within a 6 Å radius of DNA atoms. **b**, E-DNA binding site. The E-DNA (blue) is bound to the N-coil (light blue) and C-coil (red) of the Ioc3 CLB. A crystallographic two-fold axis (not shown) passes between bases T_{-1} and A_1 in the sequence $5'-T_{-1}A_1C_2G_3C_4T_5$ (complementary strand $3'-A'_{-1}T'_1G'_2C'_3G'_4A'_5$). **c**, I-DNA binding site of SANT. **d**, I-DNA binding site of SLIDE. In **b–d**, amino acids potentially important for DNA interaction are labelled. Possible hydrogen bonds are black. Hydrophobic contacts are orange.

hydrogen bonds. Nevertheless, the structure suggests that E-DNA binding may have a sequence preference.

The second DNA copy (I-DNA: related to internal-linker DNA, see dinucleosome model) is curved, permitting contacts simultaneously with SANT, SLIDE and HLB (Fig. 2a). SANT and SLIDE clearly bind the DNA minor groove, not the major groove as for the structurally related, sequence-specific R2 and R3 DNA-binding domains of cMyb (refs 19, 24, 25). The iodinated DNA did not show distinct peaks for I-DNA, indicating a lack of sequence preference. The SANT N-terminal α4 lies across the minor groove with side chains K889, L890, R893 and T897 juxtaposed. K889-Nε makes a possible hydrogen bond with a phosphate oxygen, and L890 makes a hydrophobic contact with a deoxyribose C5-methyl group (Fig. 2c). In contrast, the cMyb domains have their C-terminal α-helix aligned with the major groove, with side chains making cognate interactions. The homologous SANT-α6 of HSS–Ioc3 is sandwiched between the HAND-α3 C terminus and HSSB tip.

For SLIDE, the C-terminal α11 lies nearly parallel to the minor groove, whereas for cMyb this helix lies in the major groove (Fig. 2d). The α11 side chains R1044, N1047 and K1055, and the preceding R1037, are near the DNA backbone, but only R1037 makes a possible hydrogen bond with a phosphate group. The N-terminal coil preceding α8 containing N987 and N988 approaches the DNA backbone from the major groove side. Adjacent to these SLIDE interactions, Ioc3 side chains H415, K417 and K418 are near the DNA phosphodiester chain on the opposite side of the major groove. The I-DNA is positioned such that an extension of its trajectory by one-half turn results in major groove contact with HLB-α9, the N terminus of HLB-α10, and the loop connecting them. Therefore, the I-DNA

binding site is probably composed of three successive, nonspecific DNA-binding elements.

Cryo–EM structures with mononucleosomes

To visualize ISW1a(ΔATPase) in a context nearer to chromatin, we prepared two HSS–Ioc3–mononucleosome complexes and determined their structures using cryo-EM and single-particle analysis. One mononucleosome (45N29) has 45- and 29-bp segments extending from the boundaries of the 147-bp nucleosome core DNA (NCB), and the other (45N0) has only the 45-bp extension. Both contain the '601' sequence to obtain a strongly positioned nucleosome core[26]. The electron densities for the complexes at 22 Å and 24 Å resolution, respectively, were manually fit with the X-ray structures of DNA-bound HSS–Ioc3 and the nucleosome core particle[27]. The distinctive shapes of these components strongly favour one chirality of the EM density (Fig. 3 and Supplementary Fig. 6).

For the 45N29 complex, HSS–Ioc3 is bound between densities that represent the 45- and 29-bp DNA extensions (Fig. 3a). The longer tubular density coincides with the I-DNA, and the shorter tubular density leads to the E-DNA binding site. Only the 10-bp segment of E-DNA contacting CLB and encompassed by EM density is included in the model. The trajectory of the E-DNA in the crystal structure is probably affected by packing contacts and diverges from the obvious EM density connecting a NCB with the CLB. The E-linker DNA was completed by insertion of a 19-bp DNA segment between the NCB and CLB-bound E-DNA. Ioc3 faces the centre of the nucleosome core DNA with HLB located opposite a major groove adjacent to the nucleosome core dyad. In this position, HLB would contact DNA near an NCB but is too distant to contact the nucleosome-superhelix central turn. The I-linker DNA was completed by insertion of a 5-bp DNA segment between this NCB and the I-DNA, and by the addition of an 11-bp segment to the opposite I-DNA terminus. The elongated HSS domain extends away from the nucleosome core at an angle of ~60° to the nucleosome dyad axis and lies adjacent to the plane containing the nucleosome core dyad and superhelix axes. The density surrounding HSS/I-DNA is broader in this plane than required, possibly due to rotational disorder between HSS–Ioc3 and the nucleosome core.

In the 45N0 complex, the nucleosome core is at the opposite end of the I-DNA as compared to the nucleosome core in the 45N29 complex (Fig. 3b). The EM density indicates that the 45N0 complex is a two-fold symmetric dimer of HSS–Ioc3–45N0, a consequence of the 45-bp extensions binding to both SANT–SLIDE–HLB and CLB in different monomeric units. To model the entire 45 bp of an extension

in each dimer half, a 17-bp DNA segment was inserted into the tube of density connecting the I-DNA for one HSS–Ioc3–45N0 unit with the E-DNA of the other. The overall fit of the nucleosome core and HSS–Ioc3–45N0 to the density places the DNA backbone at the dyad position in the vicinity of the Ioc3 basic loop (Supplementary Fig. 5).

Gel mobility and photocrosslinking

We used site-directed photocrosslinking in solution to show that HSS–Ioc3 binds 45N29 and 45N0 in opposed orientations at opposite ends of the I-DNA binding site, corroborating the two cryo-EM structures. Site-directed photocrosslinking was enabled by preparing individual HSS–Ioc3 cysteine replacements for multiple amino acid positions based on the HSS–Ioc3 crystal structures, and subsequent labelling with 4-azidophenacyl bromide (APB). Four versions of reagent-modified HSS–Ioc3 yielded DNA crosslinks: (1) CLB(K695C) in the E-DNA binding site, and (2) HLB(Q467C), (3) SLIDE(N987C) and (4) SANT(L890C) all in the I-DNA binding site. All versions of site-labelled HSS–Ioc3–45N29 yield a single predominant band on native PAGE corresponding to the HSS–Ioc3–nucleosome complex visualized by cryo-EM (Fig. 4a). In comparison, HSS–Ioc3–45N0 yields a predominant band migrating with approximately half the mobility of HSS–Ioc3–45N29 in all but one case, consistent with the dimeric particle visualized by cryo-EM (Fig. 4b). The exception, CLB(K695C) with a modified E-DNA binding site, has mobility comparable to the monomeric HSS–Ioc3–45N29, consistent with CLB mediating dimer formation.

After photocrosslinking, the end-labelled DNA in a complex was cleaved with base and analysed by denaturing PAGE (Fig. 4c, d and Supplementary Fig. 7). For the E-DNA binding site, CLB(K695C) maps 19–20 bp from the 45N29 NCB, primarily in the 29-bp extension. The weaker crosslink in the 45-bp extension indicates that for CLB(K695C) the difference in DNA-extension length does not limit nucleosome binding to a single orientation about its dyad. CLB(K695C) does not produce a crosslink for 45N0, evidently due to the absence of a DNA extension in the E-DNA binding site. For the I-DNA binding site, HLB(Q467C), SLIDE(N987C) and SANT(L890C) map along the 45-bp extension chiefly at −1 to 2 bp, 10–14 bp and 26–33 bp, respectively, from the 45N29 NCB, and at 21–23 bp, 12–14 bp and −2 to −5 bp, respectively, from the 45N0 NCB. Secondary crosslinks at 10-bp intervals to the primary sites for HLB(Q467C) and SLIDE(N987C) indicate less prominent translation settings occurring for the 45-bp extension. Although APB-modified cysteine has a span of ~11 Å from Cα to reactive nitrene group, these crosslinks are
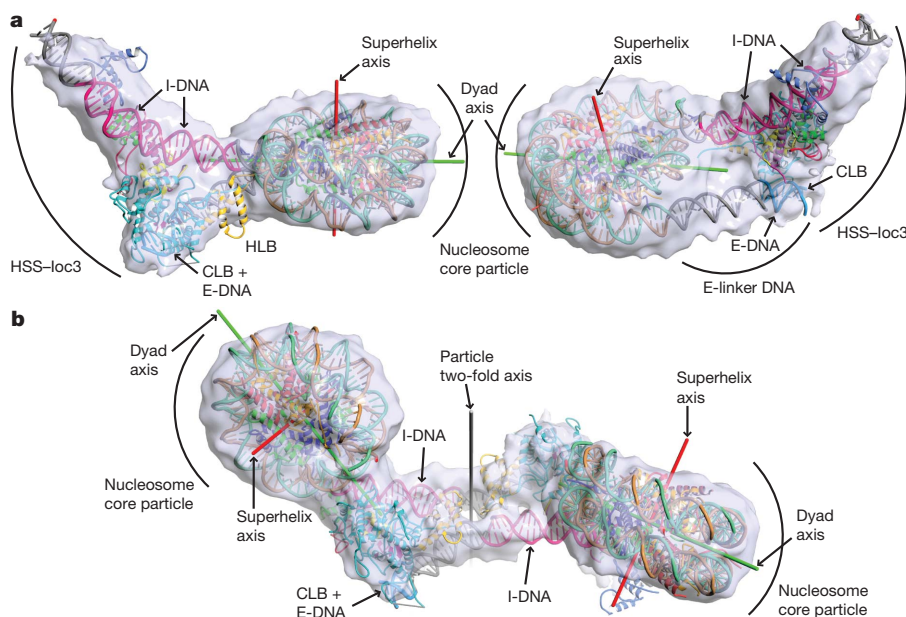

Figure 3 | ISW1a(ΔATPase)–mononucleosome cryo-EM structures. The ISW1a(ΔATPase)–DNA and nucleosome core particle X-ray structures are located in the electron densities of the two complexes. **a**, ISW1a(ΔATPase) and a nucleosome with 45- and 29-bp extensions (45N29) in two views. **b**, ISW1a(ΔATPase) and a nucleosome with a single 45-bp extension (45N0). The particle two-fold axis is shown (grey). In both parts, the nucleosome core particle superhelix (red) and dyad (green) axes are indicated. DNA segments not included in the nucleosome core particle or HSS–Ioc3 structures were added (grey).
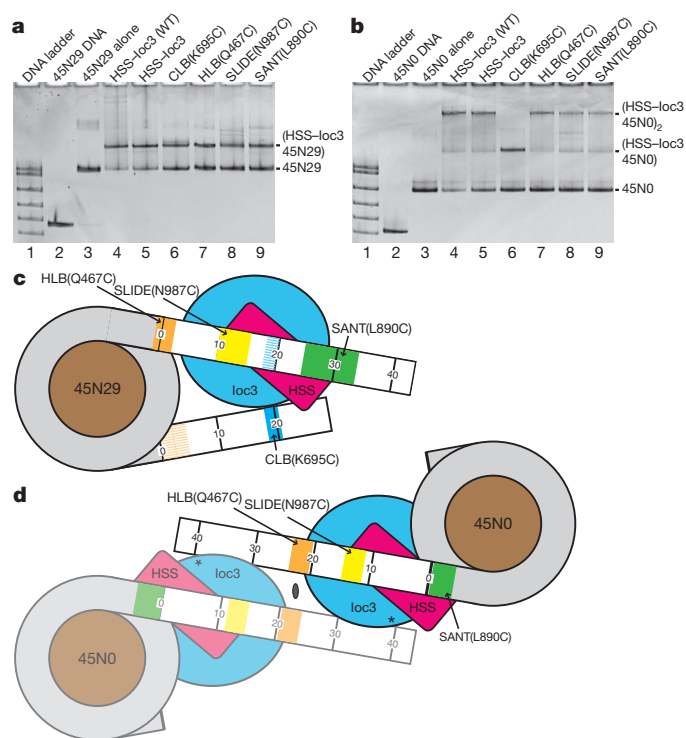
Figure 4 | Gel mobility and site-directed photocrosslinking of ISW1a(ΔATPase) bound to mononucleosomes. **a**, Native PAGE of complexes between HSS–Ioc3 and nucleosome 45N29. Lane 1, marker DNA ladder; lane 2, DNA alone; lane 3, nucleosome alone; lane 4: wild-type (WT) complex; lane 5, complex with the six native cysteines of HSS–Ioc3 replaced by alanine and used in the following lanes; lanes 6–9, complexes containing a single cysteine replacement and labelled with unphotolysed APB. **b**, Native PAGE of complexes between HSS–Ioc3 and nucleosome 45N0. The lanes are as for panel **a**. **c**, Schematic summary of photocrosslinking for HSS–Ioc3–45N29. The primary sites of crosslinking (subdomain colours as in Fig. 1) on the DNA extensions (white, with ∼10-bp intervals from the NCB marked) and corresponding APB-labelled protein locations are indicated (arrows). The weaker crosslinking sites for the complex with DNA extensions interchanged are also shown (coloured stripes). **d**, Schematic summary of photocrosslinking results for HSS–Ioc3–45N0 (labels and colours as for panel **c**). The primary sites of photocrosslinking for the mononucleosomes in the HSS–Ioc3–45N0 dimer (one half dimmed) have the opposite order from the NCB (0) compared to the 45N29 complex. The CLB(K695C)-modified complex (location marked with asterisk) is monomeric. See Supplementary Fig. 7 for the mapping data.

consistent with the close proximity of HLB(Q467C), SLIDE(N987C) and SANT(L890C) to the I-DNA in the HSS–Ioc3–DNA crystal structure and substantiate the HSS–Ioc3–45N29 and HSS–Ioc3–45N0 cryo-EM structures.

The overall crosslinking pattern for HLB(Q467C), SLIDE(N987C) and SANT(L890C) are complementary between HSS–Ioc3–45N29 and HSS–Ioc3–45N0 (Fig. 4c, d). Where the crosslinking site for 45N29 is closest to an NCB, the equivalent site for 45N0 is farthest from the same NCB, and vice versa, indicating that the 45-bp extensions for these two nucleosomes lie in opposite orientations along the I-DNA binding site. This result suggests that ISW1a would bind two successive nucleosome cores in a nucleosome array. Approximate superposition of the crosslinking patterns for HSS–Ioc3–45N29 and HSS–Ioc3–45N0 along the I-DNA binding site indicates that a common I-DNA linker would be approximately 2–3 double helical turns in length. SLIDE(N987C) has nearly a central location in the I-DNA binding site and allows the most reliable estimate of I-linker length for a hypothetically bound dinucleosome. Both HSS–Ioc3–45N29 and HSS–Ioc3–45N0 containing SLIDE(N987C) yield ∼13 bp for the distance from their NCB to the centre of the crosslinked region. Consequently, the effective I-linker length is in the vicinity of 26 bp.

Dinucleosome model and functional test

We propose that in chromatin, ISW1a interacts simultaneously with two adjoining nucleosomes to set the length of linker DNA between them. From this study, the complex containing the 45N29 nucleosome corresponds to an internal- and external-linker (I/E-linker)-bound nucleosome (I/E-nucleosome), and one monomer in the complex containing the 45N0 nucleosome corresponds to an internal-linker (I-linker)-bound nucleosome (I-nucleosome). An ISW1a(ΔATPase)–dinucleosome model was built by superposition of the common HSS–Ioc3–DNA structure contained in the HSS–Ioc3–45N29 and HSS–Ioc3–45N0 cryo-EM models (Fig. 5a). The resulting I-linker connecting NCBs of the I/E-nucleosome and I-nucleosome is 25-bp long and comprises the I-DNA external to the I-nucleosome and a 5-bp B-form DNA segment modelled to join the I-DNA to the I/E nucleosome. Potential distortion of this HLB-bound connecting segment introduces an uncertainty in I-linker length of about ±2 bp. Nevertheless, the photocrosslinked positions are highly supportive of this model (Supplementary Fig. 8). Crosslinks at secondary sites for HLB(Q467C)–45N0 suggest possible additional I-linker lengths of roughly 5 and 15 bp, but these are of insufficient length for interaction of I-linker with HLB, SLIDE and SANT simultaneously. On the basis of genomic DNA sequence periodicities, linker length is described generally by $10.2n + 5$ bp where $n = 2$ for *S. cerevisiae*[28], again indicating that the 25-bp length is most relevant.

Remodelling of mononucleosomes by ISW1a was analysed previously[29]. Translocation occurred for nucleosomes with one, but not two, 33-bp extension. Similarly, the rate of ATP hydrolysis was elevated for the nucleosome having one extension. These ISW1a–mononucleosome complexes were also footprinted with hydroxyl radical[29]. Importantly, these footprints revealed that the ATPase domain binds only the nucleosome with single extension, localized to a position 2–3 DNA turns from the dyad. In terms of our model, these mononucleosomes with one and two extensions correspond to the I- and I/E-nucleosomes, respectively, with the ATPase engaging the I-nucleosome in the half of the DNA superhelix that adjoins the I-linker (Fig. 5a).

Accordingly, we suggest two potential mechanisms for the ISW1a nucleosome spacing reaction (Fig. 5b). In both, the I/E-nucleosome would be static with its E-linker DNA bound to Ioc3 CLB, and the ATPase domain would bind and reposition the I-nucleosome. The schemes differ in which element is flexible during remodelling. (1) The I-linker is flexible and shortened until Ioc3 HLB reaches the I/E-nucleosome NCB. Flexibility may also be required of E-linker or the I/E-nucleosome DNA. I-nucleosome and the entire ISW1a translocate together. (2) The protein chain connecting HAND with the ATPase domain is flexible. The I-linker is shortened until the I-nucleosome NCB reaches SANT. Only the Isw1 ATPase translocates with the I-nucleosome. In both mechanisms, the final I-linker length is measured by the HLB–SLIDE–SANT protein-ruler. The details of translocation are not addressed here, but notably, the polypeptide connecting the ATPase domain to HAND is sufficiently long to allow the ATPase to bind in either orientation on the DNA in either scheme. Hence, I-linker length could increase or decrease even if the ATPase motor were unidirectional.

We tested our model by running a spacing reaction using a dinucleosome assembled with nucleosome cores separated by a 40-bp linker (29N40N0). I/E and I nucleosomes in this dinucleosome are defined by their positions 29 bp and 0 bp, respectively, from opposite DNA termini. The recombinant ISW1a protein used comprises full-length Ioc3 and Isw1 including the ATPase domain (B.A., and T.J.R., manuscript in preparation). After pre-incubation of the remodelling factor and 29N45N0 followed by ATP addition, the mobility of the dinucleosome in native PAGE decreases over the course of the reaction (Supplementary Fig. 9). We measured the translocation distance for both nucleosomes using site-directed, DNA mapping based on hydroxyl radical generation from a site in histone H4 (ref. 30). In this method, the primary DNA cleavages produced per strand are 2 bp
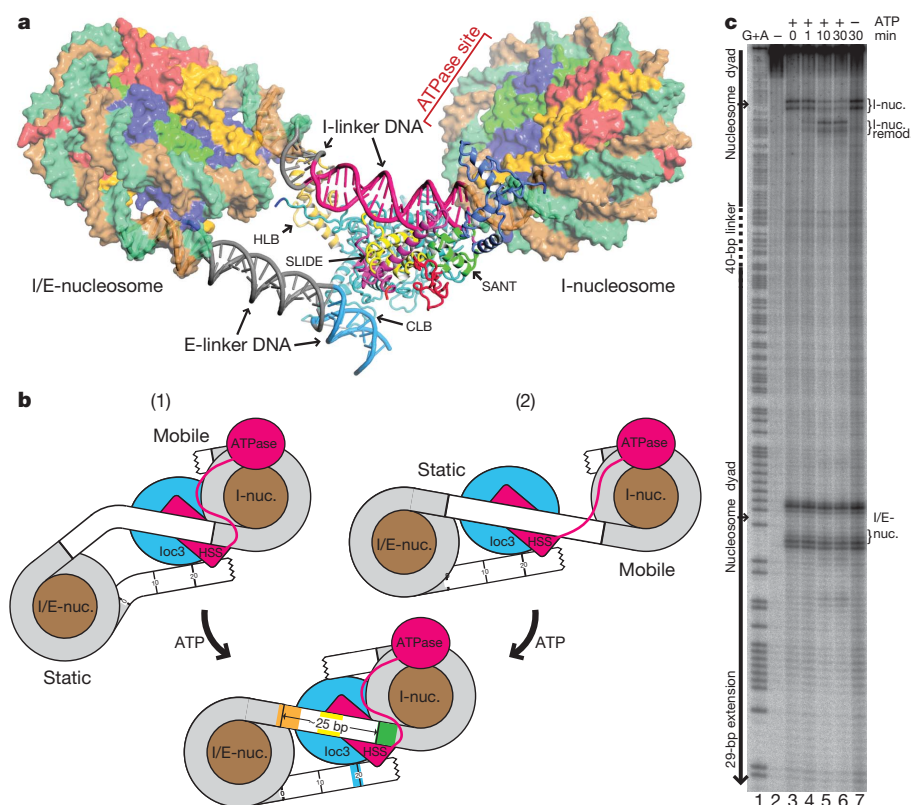
Figure 5 | Model of ISW1a(ΔATPase) bound to a dinucleosome and ISW1a remodelling reaction.
a, Overall model. A 25-bp I-linker DNA connects the two nucleosomes. HSS–Ioc3 colours are as in Fig. 1a. Segments of the I- and E-linker DNA (red and blue, respectively) are observed in the ISW1a(ΔATPase)–DNA crystal structure. Two segments (grey) were introduced to connect them to the NCB. The nucleosome core colours distinguish histones (H3, blue; H4, green; H2A, yellow; H2B, red) and DNA strands (tan and celadon). The binding site mapped for the ISW1a ATPase domain is indicated[29]. **b**, Schematics showing two mechanisms for a dinucleosome spacing reaction. In scheme (1), the linker DNA is flexible, and in scheme (2), the protein chain between HAND and the ATPase domain is flexible. **c**, ATP-dependent remodelling reaction for ISW1a with a 29N40N0 dinucleosome substrate. Denaturing PAGE. Lane 1, G+A track of 29-'601'-40-'601'-0 DNA used for nucleosome position quantification; lane 2, reaction mixture without hydroxyl radical generation; lanes 3–6, DNA footprint at 0, 1, 10 and 30 min after ATP addition; lane 7, DNA footprint at 30 min without ATP addition.

5′ and 5/6 bp 3′ to the nucleosome dyad for each different position. We observe that I/E-nucleosome maintains its starting position, and that I-nucleosome moves towards the I/E-nucleosome, reducing I-linker length to 19–20 bp (Fig. 5c). This result confirms three expectations from our model: (1) ISW1a would be oriented on this dinucleosome substrate, most probably through Ioc3 CLB interaction with the 29-bp extension; (2) the I/E-nucleosome would be static; and (3) the I-nucleosome would be mobile through association with the Isw1 ATPase. Confirmation of the ISW1a orientation awaits further studies using our ISW1a construct to assess the effect of the ATPase domain (B.A., T.D.F., K.Y. and T.J.R., manuscript in preparation). For 29N45N0, the I-nucleosome moves a one-half double helical turn farther than the 25-bp estimate from the model. The observed spacing is probably a consequence of the strong ~10-bp periodicity of the 601 sequence affecting I-nucleosome position. Stretching of I-linker and nucleosome core DNA[31–34] could effectively lengthen the linker while maintaining the stability evident for the dinucleosome complex.

Implications

ISW1a is important for transcriptional repression of many genes, and is particularly well studied for a subset of TATA-box-containing genes including *MET16* (refs 16, 35, 36). For the *MET16* promoter, both nucleosome −1, containing the TATA box and transcription initiation site, and nucleosome +1 are well positioned in the repressed state and displaced from their translational setting upon gene activation[35]. Chromatin immunoprecipitation (CHIP) analysis shows that Ioc3 is localized primarily to nucleosome −1 during repression. To establish repression, we envisage that ISW1a binds −2 and −1 successively as the I/E-nucleosome, repositioning −1 and +1, respectively, as the I-nucleosome. In both steps, the orientation of ISW1a, possibly facilitated by a DNA sequence preference imparted by Ioc3 CLB, would be important so that only the misplaced nucleosome is remodelled.

Yeast having Isw1 ATPase inactivated retain approximately sevenfold repression of *MET16* compared to yeast with Isw1 deleted entirely, indicating that ISW1a bound at nucleosome −1 interferes with binding of TBP and RNA Pol II[36]. Our results show that the ATPase domain

is not required for stable ISW1a–nucleosome binding. Deletion of SANT or SLIDE prevents nucleosome remodelling, but ΔSANT provides approximately fourfold greater repression than ΔSLIDE, suggesting that repression is due to an I/E-nucleosome. These same deletions in pull-down assays and CHIP analyses demonstrate the greater importance of SLIDE compared to SANT for Isw1–Ioc3 association, congruent with the HSS–Ioc3 structure.

Mobility studies of mononucleosomes assembled at various locations on mononucleosome-sized DNA fragments show that ISW1a remodelling results in NCB not closer than ~15 bp from a fragment end[29,37]. Our results suggest that eventual formation of an I/E-nucleosome will determine this end state via preference for binding two DNA extensions of at least 15 bp. Disruption of the E-linker DNA binding site may allow a nucleosome to reside at a DNA terminus, as for Isw1b, which contains subunits Ioc2 and Ioc4 instead of Ioc3 (ref. 37).

The HSS domain of ISWI chromatin remodelling factors combines with various proteins to impart unique functional characteristics distinguishing ISWI family members. Ioc3 evidently enables a role for Isw1 in transcriptional repression and possibly contributes to promoter recognition. Further structural and biochemical studies of ISW1a and other ISWI family members using mononucleosomes and oligonucleosomes will further illuminate the mechanism and the importance of compositional diversity in chromatin remodelling.

METHODS SUMMARY

R_{work} and R_{free} values of 28.3% and 29.7% for ISW1a(ΔATPase), and 28.3% and 29.1% for ISW1a(ΔATPase)–DNA (Supplementary Table 1a, b). Computer programs used were XDS[40], CCP4-SCALA[41], HKL2MAP[42], SHARP[43], SOLOMON[44], COOT[45] and PHENIX[46].

Vitrified samples were imaged with an F20 TEM (FEI) using a 2k × 2k CCD detector with energy filter (Gatan) for HSS–Ioc3–45N29, and a 4k × 4k CCD (Gatan) for HSS–Ioc3–45N0. Single particle analysis of HSS–Ioc3–45N29 on 13,206 particles and of HSS–Ioc3–45N0 on 4,217 particles with two-fold symmetry imposed yielded 22 Å and 24 Å resolution densities, respectively. Computer programs used were EMAN-BOXER[47] and SPIDER[48]. CHIMERA[49] was used to fit the X-ray structure to the EM density.

For photocrosslinking, cysteine replacements were introduced into HSS–Ioc3 by site-directed mutagenesis, labelled with 4-azidophenacyl bromide, and combined with either nucleosome 45N29 or 45N0 (radioactively end-labelled DNA). Samples were exposed to ultraviolet light and DNA subsequently cleaved by base and analysed by denaturing PAGE. Mapping of 29N40N0 was performed by a site-directed Fenton reaction[30].

Molecular illustrations and the dinucleosome model were prepared with PyMOL (http://www.pymol.org/) and UCSF Chimera[49].

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Lee, W. et al. A high-resolution atlas of nucleosome occupancy in yeast. Nature Genet. 39, 1235–1244 (2007).
2. Valouev, A. et al. A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Res. 18, 1051–1063 (2008).
3. Kaplan, N. et al. The DNA-encoded nucleosome organization of a eukaryotic genome. Nature 458, 362–366 (2009).
4. Sasaki, S. et al. Chromatin-associated periodicity in genetic variation downstream of transcription start sites. Science 323, 401–404 (2009).
5. Mavrich, T. N. et al. Nucleosome organization in the Drosophila genome. Nature 453, 358–362 (2008).
6. Schones, D. E. et al. Dynamic regulation of nucleosome positioning in the human genome. Cell 132, 887–898 (2008).
7. Clapier, C. R. & Cairns, B. R. The biology of chromatin remodeling complexes. Annu. Rev. Biochem. 78, 273–304 (2009).
8. Tsukiyama, T., Palmer, J., Landel, C. C., Shiloach, J. & Wu, C. Characterization of the imitation switch subfamily of ATP-dependent chromatin-remodeling factors in Saccharomyces cerevisiae. Genes Dev. 13, 686–697 (1999).
9. Vary, J. C. Jr et al. Yeast Isw1p forms two separable complexes in vivo. Mol. Cell. Biol. 23, 80–91 (2003).
10. Fyodorov, D. V. & Kadonaga, J. T. Dynamics of ATP-dependent chromatin assembly by ACF. Nature 418, 896–900 (2002).
11. Eberharter, A. et al. Acf1, the largest subunit of CHRAC, regulates ISWI-induced nucleosome remodelling. EMBO J. 20, 3781–3788 (2001).
12. Mizuguchi, G., Tsukiyama, T., Wisniewski, J. & Wu, C. Role of nucleosome remodeling factor NURF in transcriptional activation of chromatin. Mol. Cell 1, 141–150 (1997).
13. Becker, P. B. & Horz, W. ATP-dependent nucleosome remodeling. Annu. Rev. Biochem. 71, 247–273 (2002).
14. Lusser, A. & Kadonaga, J. T. Chromatin remodeling by ATP-dependent molecular machines. Bioessays 25, 1192–1200 (2003).
15. Corona, D. F. & Tamkun, J. W. Multiple roles for ISWI in transcription, chromosome organization and DNA replication. Biochim. Biophys. Acta 1677, 113–119 (2004).
16. Mellor, J. & Morillon, A. ISWI complexes in Saccharomyces cerevisiae. Biochim. Biophys. Acta 1677, 100–112 (2004).
17. Dürr, H., Korner, C., Muller, M., Hickmann, V. & Hopfner, K. P. X-ray structures of the Sulfolobus solfataricus SWI2/SNF2 ATPase core and its complex with DNA. Cell 121, 363–373 (2005).
18. Thomä, N. H. et al. Structure of the SWI2/SNF2 chromatin-remodeling domain of eukaryotic Rad54. Nature Struct. Mol. Biol. 12, 350 (2005).
19. Grüne, T. et al. Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. Mol. Cell 12, 449–460 (2003).
20. Holm, L., Kaariainen, S., Rosenstrom, P. & Schenkel, A. Searching protein structure databases with DaliLite v.3. Bioinformatics 24, 2780–2781 (2008).
21. Horton, J. R. et al. Structure of the SANT domain from the Xenopus chromatin remodeling factor ISWI. Proteins 67, 1198–1202 (2007).
22. Cuperus, G. & Shore, D. Restoration of silencing in Saccharomyces cerevisiae by tethering of a novel Sir2-interacting protein, Esc8. Genetics 162, 633–645 (2002).
23. Gavin, A. C. et al. Proteome survey reveals modularity of the yeast cell machinery. Nature 440, 631–636 (2006).
24. Ogata, K. et al. Solution structure of a specific DNA complex of the Myb DNA-binding domain with cooperative recognition helices. Cell 79, 639–648 (1994).
25. Tahirov, T. H. et al. Mechanism of c-Myb-C/EBPβ cooperation from separated sites on a promoter. Cell 108, 57–70 (2002).
26. Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J. Mol. Biol. 276, 19–42 (1998).
27. Davey, C. A., Sargent, D. F., Luger, K., Mäder, A. W. & Richmond, T. J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. J. Mol. Biol. 319, 1097–1113 (2002).
28. Wang, J. P. et al. Preferentially quantized linker DNA lengths in Saccharomyces cerevisiae. PLOS Comput. Biol. 4, e1000175 (2008).
29. Gangaraju, V. K. & Bartholomew, B. Dependency of ISW1a chromatin remodeling on extranucleosomal DNA. Mol. Cell. Biol. 27, 3217–3225 (2007).
30. Flaus, A., Luger, K., Tan, S. & Richmond, T. J. Mapping nucleosome position at single base-pair resolution by using site-directed hydroxyl radicals. Proc. Natl Acad. Sci. USA 93, 1370–1375 (1996).
31. Luger, K., Maeder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 389, 251–260 (1997).
32. Makde, R. D., England, J. R., Yennawar, H. P. & Tan, S. Structure of RCC1 chromatin factor bound to the nucleosome core particle. Nature 467, 562–566 (2010).
33. Ong, M. S., Richmond, T. J. & Davey, C. A. DNA stretching and extreme kinking in the nucleosome core. J. Mol. Biol. 368, 1067–1074 (2007).
34. Vasudevan, D., Chua, E. Y. D. & Davey, C. A. Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. J. Mol. Biol. 403, 1–10 (2010).
35. Morillon, A. et al. Isw1 chromatin remodeling ATPase coordinates transcription elongation and termination by RNA polymerase II. Cell 115, 425–435 (2003).
36. Pinskaya, M., Nair, A., Clynes, D., Morillon, A. & Mellor, J. Nucleosome remodeling and transcriptional repression are distinct functions of Isw1 in Saccharomyces cerevisiae. Mol. Cell. Biol. 29, 2419–2430 (2009).
37. Stockdale, C., Flaus, A., Ferreira, H. & Owen-Hughes, T. Analysis of nucleosome repositioning by yeast ISWI and Chd1 chromatin remodeling complexes. J. Biol. Chem. 281, 16279–16288 (2006).
38. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292, 195–202 (1999).
39. Berger, I., Fitzgerald, D. J. & Richmond, T. J. Baculovirus expression system for heterologous multiprotein complexes. Nature Biotechnol. 22, 1583–1587 (2004).
40. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Cryst. 26, 795–800 (1993).
41. Evans, P. Scaling and assessment of data quality. Acta Crystallogr. D 62, 72–82 (2006).
42. Pape, T. & Schneider, T. R. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. J. Appl. Cryst. 37, 843–844 (2004).
43. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. Acta Crystallogr. D 59, 2023–2030 (2003).
44. Abrahams, J. P. & Leslie, A. G. Methods used in the structure determination of bovine mitochondrial F1 ATPase. Acta Crystallogr. D 52, 30–42 (1996).
45. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. Acta Crystallogr. D 60, 2126–2132 (2004).
46. Adams, P. D. et al. PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr. D 58, 1948–1954 (2002).
47. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: Semiautomated software for high-resolution single-particle reconstructions. J. Struct. Biol. 128, 82–97 (1999).
48. Frank, J. et al. SPIDER and WEB: Processing and visualization of images in 3D electron microscopy and related fields. J. Struct. Biol. 116, 190–199 (1996).
49. Pettersen, E. F. et al. UCSF Chimera - A visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612 (2004).

Author Contributions K.Y. was responsible for X-ray crystallography. T.D.F. was responsible for electron microscopy. B.A. was responsible for mobility assays, site-directed photocrosslinking and hydroxyl radical footprinting. D.J.F. and C.D. designed and expressed the first constructs of ISW1a(ΔATPase) and carried out biochemical analyses. K.Y. and B.A. designed the specific ISW1a constructs essential to the X-ray crystallography and electron microscopy, and to the photocrosslinking and footprinting, respectively. K.S. provided biochemical technical assistance. D.F.S. provided assistance for X-ray data collection. T.J.R. conceived the project, supervised and constructed the dinucleosome model. T.J.R. prepared the manuscript with major contributions from K.Y., T.D.F. and B.A. All authors discussed the results and commented on the manuscript.

Author Information X-ray structures have been deposited in the Protein Data Bank under accession numbers 2y9y (ISW1a(ΔATPase)) and 2y9z (ISW1a(ΔATPase)–DNA), respectively. Cryo-EM maps have been deposited in the Electron Microscopy Data Bank under accession codes 1877 and 1878. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be address to T.J.R. (richmond@mol.biol.ethz.ch).

METHODS

ISW1a(ΔATPase) preparation. The Isw1 and Ioc3 sequences were analysed by limited proteolysis and secondary structure prediction[38] (K.Y. and T.J.R., unpublished data), and subsequently Isw1 amino acids 763–1129 (HSS-His$_6$) and Ioc3 amino acids 127–749 were co-expressed using the MultiBac insect cell system following the protocol of ref. 50. Harvested SF21 cells were lysed in 50 mM TrisHCl, pH 7.0, 100 mM KCl, 0.1% Nonidet P-40, 1.3 µg ml^{-1} leupeptin, 1.3 µM pepstatin A, 0.13 mg ml^{-1} PMSF, 1:750 protease inhibitor cocktail (Sigma) and fractionated into cytosolic and nuclear components. The nuclear pellet was soaked in 50 mM TrisHCl, pH 7.0, 400 mM KCl, 1.3 µM pepstatin A, 0.13 mg ml^{-1} PMSF, 1:750 protease inhibitor cocktail (Sigma) for 1–2 h at 4 °C. After removing nuclear debris, the ISW1a(ΔATPase) in the supernatant was purified by chromatography on Talon (Clontech) using 50 mM TrisHCl, pH 7.0, 400 mM NaCl, 200 mM imidazole for elution, and on HiTrap heparin (GE Healthcare) using a linear gradient of 50 mM to 1 M NaCl in 20 mM TrisHCl, pH 7.0, 1 mM EDTA, 1 mM DTT, 0.05% β-mercaptoethanol and 10% glycerol for elution. Effluent after concentration (Amicon Ultra-4, Millipore) underwent gel filtration using Superose 6 (GE Healthcare) either in 20 mM TrisHCl, pH 7.0, 400 mM NaCl, 1 mM EDTA, 1 mM DTT, 0.05% β-mercaptoethanol, 10% glycerol for the protein alone or in the same solution with 100 mM NaCl for the protein–DNA complex. The peak fractions were concentrated to 25 mg ml^{-1} (Amicon Ultra-4, Millipore). AKTA-purification systems (GE Healthcare) were used for column chromatography. All oligonucleotides used for cloning were purchased from Microsynth AG. The six native cysteines in HSS and Ioc3 were changed to alanine for photocrosslinking experiments.

Histone octamer and nucleosome preparation. Recombinant *Xenopus laevis* histones were expressed in bacteria, purified and assembled into histone octamers and mononucleosomes or dinucleosomes as described[51]. Histone mutants H4(S47C) and H3(C110A) were used for mapping with hydroxyl radical. Mononucleosomes at 5 µM were dialysed into 10 mM TrisHCl, pH 7.5, 10 mM KCl. Dinucleosomes were purified by PAGE (5% acrylamide:bisacrylamide 37.5:1) in 22.3 mM Tris, 22.2 mM boric acid, 2.5 mM EDTA at 20 °C with circulation of the electrolyte solution. After exposure of the gel to film (Fuji Super RX), gel pieces containing dinucleosome were excised and eluted into 10 mM TrisHCl, pH 7.5, 10 mM KCl, 10% glycerol, 0.5 mg ml^{-1} BSA at 4 °C for 12 h. The eluant was concentrated (Vivaspin 500, Sartorius Stedim) and washed with 10 mM TrisHCl, pH 7.5, 10 mM KCl. Nucleosome concentrations were determined by ultraviolet spectrometry (NanoDrop 2000, Thermo Scientific).

DNA preparation. The mononucleosome DNA used for cryo-EM and site-directed ultraviolet crosslinking experiments contained the 601 nucleosome positioning sequence[26] with either two extensions of 45 bp and 29 bp or a single extension of 45 bp, each flanked by an EcoRV restriction site (45-bp sequence, ATCGAAGACAGAGTACTTATGTGATGGACCCTATACGCGGCCGCC-601; 29-bp sequence, 601-GCATGTATTGACAGCGACCTGTCGAGGAT). Dinucleosome DNA contained two 601 nucleosome-positioning sequences, a 40-bp internal and a 29-bp external linker (40-bp sequence, GCATGTATTGAACAGCGACCTTGCCGGTGCCAGTTCTAGA; 29-bp sequence, 601-GCATGTATTGAACAGCGCTAGGCTCCGAT). All fragments were obtained from digested plasmid DNA. For radioactive labelling mononucleosomes, DNA ends were modified to XmaI and EcoRI restriction sites to permit strand-specific labelling with [α-^{32}P]dCTP using Klenow polymerase. For dinucleosomes, only the XmaI site was introduced.

ISW1a(ΔATPase) crystallography. The following alterations to ISW1a(ΔATPase) were introduced to improve the crystal diffraction quality: HSS K815Q, D844K, E848Q, K851Q, Q853E and Ioc3 N682K. Crystals were grown at 22 °C by hanging-drop/vapour diffusion using a reservoir solution of 100 mM BisTris, pH 7.0, 1.6 M Na-citrate, 10% glycerol, 0.1 M NaPO$_4$ and drops of protein stock solution diluted 1.35-fold with reservoir solution. Crystals were transferred to a solution of 50 mM BisTris, pH 6.5, 0.9 M Na-citrate, 0.1 M NaPO$_4$, 10% glycerol and 0.05% β-mercaptoethanol, followed by stepwise transfer into the same solution but with 2%, 4%, 6%, 8% and 10% ethylene glycol, cooled passively to −20 °C over 3 h, and flash-cooled into liquid N$_2$. Altogether, these procedures improved the crystalline diffraction from ~7 to 3.1 Å resolution. Two heavy-atom derivatives were prepared at 4 °C by soaking crystals in 50 mM BisTris, pH 6.5, 0.9 M Na-citrate, 0.1 M NaPO$_4$, 10% glycerol and 10% ethylene glycol containing either 1 mM methylmercury nitrate or Ta$_6$Br$_{12}$ for 10 h, followed by back-soaking without heavy atom for 3 h. Selenomethionine-labelled crystals were also prepared[52]. Slow-cooling and flash-cooling were performed as for the native crystals. Diffraction data were measured on the X06SA beamline of the Swiss Light Source at 100 K using 0.05° rotations and the Pilatus detector, and processed with XDS[40] and CCP4-SCALA[41] (Supplementary Table 1a).

Initial phases were calculated by the single-wavelength anomalous diffraction (SAD) method with HKL2MAP[42] from the methylmercury derivative data set and used to find the heavy-atom positions of the Ta$_6$Br$_{12}$ derivative by difference

Fourier maps. Two-derivative phases were used to find the selenium atom sites of the selenomethionine-containing crystals by difference Fourier maps. The structure was solved at 3.25 Å using SHARP[43] to refine heavy atom parameters for multiple isomorphous replacement and anomalous scattering (MIRAS) and SOLOMON[44] to improve phases through density modification.

The model was built with COOT[45] and refined with PHENIX[46] applying anisotropic scaling and bulk solvent correction during rigid body, multiple translation-libration-screw rotation (TLS) refinement, coordinate and individual restrained B-factor refinement, and simulated annealing. The final R_{work} and R_{free} values were 28.3% and 29.7%, respectively (Supplementary Table 1a).

ISW1a(ΔATPase)–DNA crystallography. Two DNA oligonucleotides with sequences GCGCATGAACCCGTATATAAGC and CTAGGCTTATATACG GGTTCATGCGC were heated to 98 °C and cooled to 20 °C over 10 h, followed by ligation for 10 h using T4 DNA ligase at 17 °C. The resulting palindromic 48-bp DNA duplex was purified using Hi-trapQ (GE Healthcare) chromatography with a linear gradient of 50 mM NaCl to 1 M NaCl in 20 mM TrisHCl, pH 7.0 and 1 mM EDTA, followed by dialysis to 70 mM NaCl, 20 mM TrisHCl, pH 7.0 and 1 mM EDTA. Two different iodine-labelled 48-bp DNA duplexes were prepared by combining GCGCATGAACCCGTATATAAGC with either CTAGGC T(5IdT)ATATACGGGTTCATGCGC or CTAGGCTTATA(5IdT)ACGGGTT CATGCGC following the method described above. All oligonucleotides used for crystallography were purchased from Microsynth AG. Crystals were grown at 22 °C by hanging-drop/vapour diffusion using a reservoir solution of 50 mM BisTris, pH 7.0, 150 mM Na-citrate, 30% (W/V) PEG5000, and drops of equal volumes of protein (200 µM) and DNA (400 µM) stock solutions were diluted 1.25-fold with reservoir solution. Crystals were transferred at 4 °C to a solution of 20 mM BisTris, pH 7.0, 100 mM Na-citrate, 20 mM MgCl$_2$, 0.05% β-mercaptoethanol, 10% glycerol and 10%(v/v) PEG2000 followed by stepwise transfer into the same solution but with 12%, 14%, 16% and 18% PEG2000 followed by further transfer to 30% glycerol in 4% increments. Heavy atom derivatives for methylmercury nitrate and Ta$_6$Br$_{12}$ were prepared at 4 °C in this solution without β-mercaptoethanol by soaking crystals for 10 h in 1 mM heavy atom compound, followed by back-soaking without heavy atom for 3 h. Crystals were cooled passively to −20 °C, flash-cooled in liquid propane at −95 °C, and stored in propane under liquid N$_2$. Diffraction data were measured and processed as for the crystals of the protein alone (Supplementary Table 1b).

The structure of the ISW1a(ΔATPase)–DNA complex was solved at 3.6 Å using a Ta$_6$Br$_{12}$ derivative to calculate initial phases by the SAD method with HKL2MAP[42], followed by multiple isomorphous replacement with anomalous scattering using methylmercury, Ta$_6$Br$_{12}$ and the 5Id30T and 5Id34T DNA derivatives. Model building using COOT[45] was initiated by manually fitting the ISW1a(ΔATPase) structure to the electron density and refinement was performed with PHENIX[12]. The final R_{work} and R_{free} values were 28.3% and 29.1%, respectively (Supplementary Table 1b).

Molecular illustrations were prepared with PyMOL (http://www.pymol.org/) and UCSF CHIMERA[49].

ISW1a(ΔATPase)–mononucleosome electron microscopy. ISW1a(ΔATPase)–nucleosome complexes were applied to C-flat grids (Protochips) and vitrified in liquid ethane using a Vitrobot (FEI). Electron micrographs were collected using the minimum dose technique (≤15 e Å^{-2}) on an F20 TEM (FEI) operating at −180 °C and 200 kV. HSS–Ioc3–45N29 was imaged at ×54,000 magnification using a 2k × 2k Ultrascan CCD camera (Gatan) sampling at 2.70 Å per pixel and fitted with a zero-loss energy filter (GIF tridiem, Gatan). HSS–Ioc3–45N0 was imaged at ×107,520 magnification using a 4k × 4k CCD camera (Gatan) sampling at 1.39 Å per pixel, and images were later binned to 2.78 Å per pixel. Particles were picked semi-automatically with EMAN-BOXER[47] and numbered 13,206 for HSS–Ioc3–45N29 and 4,217 for HSS–Ioc3–45N0. Particles underwent five iterations of correspondence analysis and multireference alignment using SPIDER[48]. The resulting class averages were processed using the common-lines algorithm to generate an initial three-dimensional template[53]. Three-dimensional reconstruction was performed by projection-matching refinement using SPIDER[48]. Two-fold symmetry was imposed for HSS–Ioc3–45N0. The contrast-transfer functions of the class averages were corrected by Weiner filtration. The final reconstructions for HSS–Ioc3–45N29 and HSS–Ioc3–45N0 were 22 Å and 24 Å resolution, respectively, as estimated by 0.5 Fourier shell correlation[54].

The EM electron densities were fit with the X-ray structure of the nucleosome core particle[31] (Protein Data Bank 1kx5) and the X-ray structure of ISW1a(ΔATPase)–DNA reported here using the 'Fit in Map' command of CHIMERA[49] after initial placements were made manually. For HSS–Ioc3–45N29, the terminus of the I-DNA closest the NCB was positioned and maintained in the electron density connecting them. The nucleosome core was rotated a few degrees around its superhelix axis so that a model of B-form DNA could be inserted between the I-DNA terminus and NCB. A length of 5 bp correctly joins the termini of opposing phosphodiester chains. The other terminus of I-DNA was extended

with an 11-bp B-form model to complete the 45-bp extension. The E-DNA was connected correctly to the other NCB with a 19-bp B-form model, and truncated to yield the 29-bp extension. After rigid-body refinement of the entire complex using CHIMERA[49], the correlation coefficient between the EM electron density (values over $+1\sigma$ from mean) and the calculated model density is 0.86.

For HSS–Ioc3–45N0, the fitting proceeded as for HSS–Ioc3–45N29. The region of overlap between the I-nucleosome NCB and I-DNA were optimized by manually rotating the I-nucleosome a few degrees around its superhelix axis and translating HSS–Ioc3–DNA several Ångstroms to align the phosphodiester chains at the NCB. The opposite end of the I-DNA from one monomeric unit was connected to the E-DNA bound by the two-fold-related monomer with a 16-bp B-form model having a 60 Å radius of curvature. After rigid-body refinement of the entire dimeric complex using CHIMERA[49], the correlation coefficient between the EM electron density (values over $+1\sigma$ from mean) and the calculated model density is 0.87.

The ISW1a(ΔATPase)–dinucleosome model was built with CHIMERA[49] by superimposing the common HSS–Ioc3–DNA component of the composite X-ray/cryo-EM structures (HSS–Ioc3–45N29 and one-half HSS–Ioc3–45N0).

Site-directed photocrosslinking. ISW1a(ΔATPase) was treated for 1 h with 10 mM DTT in 20 mM TrisHCl, pH 7.0, 400 mM KCl and 10% glycerol, dialysed into 20 mM TrisHCl, pH 7.5, 400 mM KCl and 10% glycerol, and reacted with a tenfold excess of 4-azidophenacyl bromide (Sigma) for 3 h in the dark at 20 °C. Free reagent was removed by dialysis against 20 mM TrisHCl, pH 7.0, 400 mM KCl and 10% glycerol. Protein concentrations were determined by Bradford assay (Bio-Rad) and protein modification was confirmed by MS. Binding reactions containing 1 µM radioactive nucleosomes and 1–1.2 µM reagent-labelled ISW1a(ΔATPase) in 25 mM TrisHCl pH 7.3, 70 mM KCl were incubated for 30 min in the dark at 20 °C, cooled to 4 °C and irradiated for 1–5 min with a UV-lamp (CAMAG) at 254 nm and 10 cm distance. Native samples were analysed by PAGE (5% acrylamide:bisacrylamide 37.5:1) in 22.3 mM Tris base, 22.2 mM boric acid, 2.5 mM EDTA at 4 °C with circulation of the electrolyte solution. Samples were denatured for 20 min at 70 °C by incubating with 1 vol. of 200 mM TrisHCl, pH 8.0 and 0.2% SDS, and followed by extraction with 1 vol. phenol/chloroform (4:1). The organic phase was washed twice with 1 vol. of 1 M TrisHCl, pH 8.0 and 1% SDS, and the crosslinked DNA was ethanol precipitated and cleaved for 30 min at 90 °C with 1 M piperidine (Fluka)[55]. After ethanol precipitation, the DNA fragments were solubilized in HiDi formamide (Applied Biosystems) and analysed by denaturing PAGE (5% or 8% acrylamide:-bisacrylamide, 19:1) using 42% urea. Band positions were measured relative to the well positions for the 5% gels.

Site-directed hydroxyl radical mapping of remodelling reactions. Octamer was modified with 100-fold excess N-[S-(2-pyridylthio)-cysteaminyl]ethylenedia-mine-N,N,N′,N′-tetraacetic acid (Toronto Research Chemicals Inc.) as decribed[30], except MES was substituted for cacodylate. Protein concentration was measured by Bradford Assay (Bio-Rad). Remodelling reactions of 130 nM 29N40N0, 10 nM ISW1a (His$_6$-Isw1ΔN$_{105}$ + Ioc3), 50 mM TrisHCl, pH 7.5, 50 mM KCl, 1 mM MgCl$_2$, 390 nM (NH$_4$)$_2$Fe(SO$_4$)$_2$·6H$_2$O (Fluka) were initiated by the addition of ATP (AppliChem) to 0.1 mM. Reactions were incubated at 30 °C before stopping by addition of 0.25 volumes apyrase (NEB, ~10 mU µl^{-1} after removal of DTT and EDTA) and incubation for 5 min. Samples were chilled on ice before 4 µl was removed for analysis by native PAGE as described above. The hydroxyl radical reaction was initiated through addition of 0.5 volumes of 12 mM ascorbic acid (Fluka) in 50 mM TrisHCl, pH 7.5, 100 mM KCl and 0.5 volumes of 0.2% H$_2$O$_2$ (Fluka). The reactions were allowed to proceed for 30 min at 20 °C. Samples were subjected to phenol/chloroform extraction and ethanol precipitation. DNA was re-suspended in 5 µl HiDi formamide (Applied Biosystems) and analysed by denaturing PAGE (4–6% acrylamide:bisacrylamide, 19:1) using 42% urea.

50. Bieniossek, C., Richmond, T. J. & Berger, I. in *Curr. Prot. Protein Sci.* 5.20.21–5.20.26 (John Wiley & Sons, 2008).
51. Luger, K., Rechsteiner, T. J. & Richmond, T. J. Expression and purification of recombinant histones and nucleosome reconstitution. *Methods Mol. Biol.* **119,** 1–16 (1999).
52. Bellizzi, J. J., Widom, J., Kemp, C. W. & Clardy, J. Producing selenomethionine-labeled proteins with a baculovirus expression vector system. *Structure* **7,** R263–R267 (1999).
53. Van Heel, M. Angular reconstitution: A posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy* **21,** 111–123 (1987).
54. Unser, M. *et al.* Spectral signal-to-noise ratio and resolution assessment of 3D reconstructions. *J. Struct. Biol.* **149,** 243–255 (2005).
55. Hu, D., Crist, M., Duan, X., Quiocho, F. A. & Gimble, F. S. Probing the structure of the PI-SceI-DNA complex by affinity cleavage and affinity photocross-linking. *J. Biol. Chem.* **275,** 2705–2712 (2000).

Imprints of fast-rotating massive stars in the Galactic Bulge

Cristina Chiappini[1,2,3], Urs Frischknecht[4,5], Georges Meynet[2], Raphael Hirschi[5,6], Beatriz Barbuy[7], Marco Pignatari[4], Thibaut Decressin[2] & André Maeder[2]

The first stars that formed after the Big Bang were probably massive[1], and they provided the Universe with the first elements heavier than helium ('metals'), which were incorporated into low-mass stars that have survived to the present[2,3]. Eight stars in the oldest globular cluster in the Galaxy, NGC 6522, were found to have surface abundances consistent with the gas from which they formed being enriched by massive stars[4] (that is, with higher α-element/Fe and Eu/Fe ratios than those of the Sun). However, the same stars have anomalously high abundances of Ba and La with respect to Fe[4], which usually arises through nucleosynthesis in low-mass stars[5] (via the slow-neutron-capture process, or s-process). Recent theory suggests that metal-poor fast-rotating massive stars are able to boost the s-process yields by up to four orders of magnitude[6], which might provide a solution to this contradiction. Here we report a reanalysis of the earlier spectra, which reveals that Y and Sr are also overabundant with respect to Fe, showing a large scatter similar to that observed in extremely metal-poor stars[7], whereas C abundances are not enhanced. This pattern is best explained as originating in metal-poor fast-rotating massive stars, which might point to a common property of the first stellar generations and even of the 'first stars'.

NGC 6522 has been confirmed to be older than any halo globular cluster, despite its metallicity being a tenth that of the Sun[4], and is therefore a witness of the early phases of the chemical enrichment of the Universe. Consistent with the age of this cluster, its stars show a chemical pattern typical of an interstellar medium enriched by core-collapse supernovae (in which thermonuclear supernovae of type Ia and low- and intermediate-mass stars did not have time to contribute to the chemical enrichment). However, the large [Ba/Eu] ratios found[4] in five of the eight stars of NGC 6522 studied (Table 1) shows that the excess in Ba cannot be attributed to the rapid-neutron-capture (r) process, and so the s-process must be invoked[4,5] (see Supplementary Information).

There are only two ways to explain the high Ba and La found in NGC 6522, namely: (1) the original gas from which the globular cluster formed had been pre-enriched in s-process elements by previous generations of massive stars, or (2) the original composition of the stars formed in the globular cluster was lately modified by mass-transfer episodes taking place in binary systems involving low-mass asymptotic giant branch (AGB) stars within NGC 6522.

From re-inspection of the spectra of NGC 6522 (ref. 4), we were able to obtain the Y abundances for eight giant stars, and estimate the Sr abundances for six of them (Table 1 and the Supplementary Information). We find large overabundances of Y (and Sr) with respect to Fe and Ba in the NGC 6522 stars, with a similar scatter to that observed in extremely metal-poor halo stars[7], but now also observed for bulge stars with [Fe/H] = −1, a result not seen previously (see Fig. 1). In addition, from the C_2 band-head (see Supplementary Information), we were able to estimate upper limits for the [C/Fe] ratio. We found that all studied stars have [C/Fe] ≲ 0.0, and hence are not enriched in C, as is the case in a significant fraction of very-metal-poor halo stars[8].

Extremely metal-poor environments can produce noticeable effects on the properties of massive stars (more details can be found in the Supplementary Information). At very low metallicities, stars rotate faster[9]. Models of fast-rotating massive stars (hereafter 'spinstars') at very low metallicities[10–12] have shown that rotational mixing transports ^{12}C from He-burning core into H-rich layers where it is transformed to ^{14}N and ^{13}C. This primary ^{14}N is then transported back to the He-burning core where it is converted into ^{22}Ne, the main neutron source in massive stars for the s-process beyond Fe. Hence, the amount of ^{22}Ne and of s-process products in the He core is enhanced with respect to non-rotating models[6].

We have calculated extremely metal-poor 'spinstar' models[13] ([Fe/H] = −3.8) with a reaction network including 613 isotopes up to Bi (U.F. *et al.*, manuscript in preparation). Rotational mixing

Table 1 | Abundances of the eight stars in NGC 6522.

Element	Reference	B-8 (star 1)	B-107 (star 2)	B-108 (star 3)	B-118 (star 4)	B-122 (star 5)	B-128 (star 6)	B-130 (star 7)	F-121 (star 8)
[O/Fe]	4	+0.25	+0.50	+0.70	+0.30	+0.70	−	+0.50	+0.50
[Mg/Fe]	4	+0.10	+0.27	+0.33	+0.20	+0.20	+0.25	+0.40	+0.40
[Si/Fe]	4	+0.34	+0.20	+0.20	+0.29	+0.13	+0.24	+0.35	+0.27
[Ca/Fe]	4	+0.15	+0.04	+0.18	+0.21	+0.21	+0.16	+0.23	+0.16
[Ti/Fe]	4	+0.12	+0.14	+0.21	+0.11	+0.19	+0.17	+0.21	+0.16
[Ba/Fe]	4	+0.95	+0.50	0.00	+1.00	+0.60	+0.90	+0.25	−0.25
[La/Fe]	4	+0.50	+0.50	+0.30	+0.50	+0.30	−	−	0.00
[Y/Fe]	**This work**	**+1.20**	**+1.30**	**+1.00**	**+0.50**	**+1.20**	**+1.50**	**+1.20**	**+1.20**
[Sr/Fe]	**This work**	**+1.20**	**+1.00**	**+1.55**	**+1.50**	**+0.50**	**+1.50**	**−**	**−**
[Eu/Fe]	4	+0.50	0.00	+0.50	+0.50	+0.30	0.00	+0.80	+0.50
[Na/Fe]	4	+0.35	−0.30	−0.15	+0.10	+0.15	+0.10	+0.15	−0.10
[C/Fe]	This work	≲0	≲0	≲0	≲0	≲0	≲0	≲0	≲0

We present the abundances reported in the literature[4] and the new [Y/Fe] and [Sr/Fe] abundances (shown in boldface) obtained here (where [A/B] = $\log(N_A/N_B) - \log(N_A/N_B)_\odot$ for the number N of atoms of elements A and B). The uncertainties in [X/Fe] = Mg, Si, Ca, Ti, Ba, Eu and Na are of 0.2 dex. The [O/Fe] ratios have larger uncertainties[4] (about 0.3 dex). The [Sr/Fe] and [La/Fe] abundances have larger uncertainties (about 0.3 dex) owing to the weakness of the available lines, and could be estimated only for six of the eight stars. Although the Sr I lines are weak and subject to non-LTE (local thermodynamic equilibrium) effects, a clear Y II line at 6,613.733 Å was measured, leading to [Y/Fe] ratios with uncertainties around 0.15 dex (see Supplementary Information). In addition, from the C_2 band-head at wavelength λ = 6,675.90 Å we were able to estimate upper limits for the [C/Fe] ratios and found that all studied stars in NGC 6522 have [C/Fe] ≲ 0.0. Hence, these stars are not strongly C-enhanced relative to Fe, as are many of the halo stars with [Fe/H] < −3.0.

[1]Astrophysikalisches Institut Potsdam, An der Sternwarte 16, Potsdam, 14482, Germany. [2]Geneva Observatory, University of Geneva, 51 Ch. des Maillettes, Sauverny, 1290, Switzerland. [3]Istituto Nazionale di Astrofisica, Osservatorio Astronomico di Trieste, Via G. B. Tiepolo 11, Trieste, 34143, Italy. [4]Department of Physics, University of Basel, Klingelbergstrasse 82, Basel, 4056, Switzerland. [5]Astrophysics Group, Keele University, ST5 5BG, Keele, England. [6]IPMU, University of Tokyo, Kashiwa, Chiba, 277-8582, Japan. [7]University of São Paulo, IAG, Rua do Matão 1226, Cidade Universitaria, 05508-900, São Paulo, Brazil.
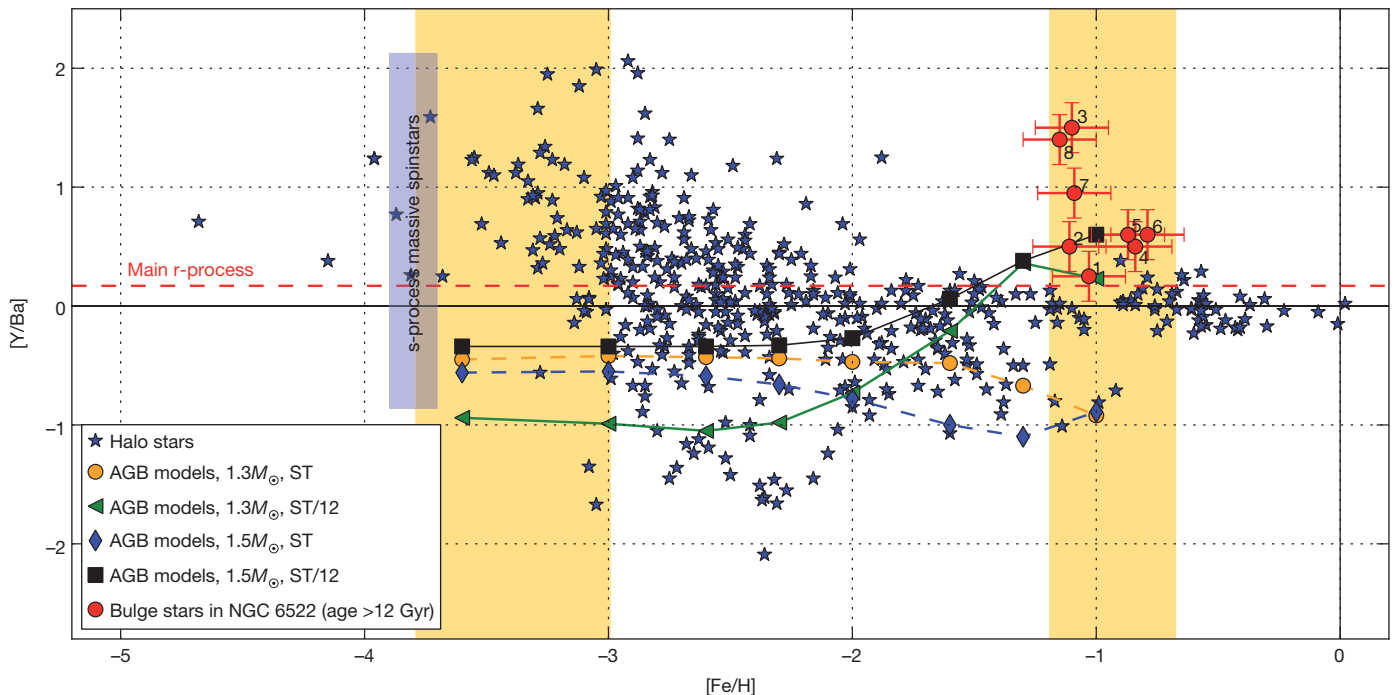
Figure 1 | The [Y/Ba] scatter observed in the early Universe. Observed [Y/Ba] scatter in the NGC 6522 stars, which have a metallicity of around [Fe/H] = −1 (ref. 4) (red circles with s.d. error bars) compared to that observed in extremely metal-poor halo stars[7] (with [Fe/H] < −3). The two yellow shaded areas mark the 'early Universe phase' sampled by halo and Bulge stars. We note the similar scatter in [Y/Ba] between the most metal-poor halo stars and our Bulge stars (compare the scatter inside the two yellow zones). Also shown is the [Y/Ba] scatter predicted for the very earliest phases of the chemical enrichment owing to 'spinstars' (indicated by the blue column). The dashed line indicates the [Y/Ba] ratio predicted from pure r-process[5]. Finally, the curves show the predictions for the [Y/Ba] ratio by recent AGB models[14], at different metallicities.

The lowest ^{13}C-pocket efficiency model considered here[14] is 'ST/12' (where ST is for standard, and ST/12 means the efficiency of the standard case reduced by a factor of 12; the ^{13}C-pocket is a tiny radiative layer of material in the He-rich region just below the H shell, responsible for most of the s-process in AGB stars). We do not consider AGB models with lower efficiencies (showing higher [Y/Ba] ratios) because the corresponding [Ba/Fe] and [Y/Fe] ratios would be too low[14] (approximately solar) compared to the large enhancements observed in the NGC 6522 stars (Table 1). Finally, we note that whereas AGB mass-transfer and 'spinstar' models can explain stars with slightly sub-solar [Y/Ba] ratios (−1 < [Y/Ba] < 0), both scenarios have difficulties in explaining stars with [Y/Ba] ratios below −1. Gyr, billion year; $M_\odot$, solar mass.

increases the s-process yields by about four orders of magnitude (Fig. 2a and b; see also Supplementary Information). The efficiency of this process depends on the rotation rate. A lower rotation rate leads to a more efficient production of the Sr peak compared to heavier species (for instance, in our models the upper limit of [Y/Ba] is about +2), whereas strong mixing driven by a high rotation rate boosts the peak of s-process products towards heavier elements, decreasing the [Y/Ba] ratio (we obtain a lower limit for [Y/Ba] of around −1), even producing non-negligible quantities of Pb. Hence, one of the main predictions of our 'spinstar' scenario is that the early Universe composition should exhibit not only a large scatter in several [s-process element/Fe] ratios, but also a large scatter in the abundance ratios of elements belonging to the different s-process peaks.

Figure 1 compares the scatter in the [Y/Ba] ratio predicted by models of 'spinstars' in the very earliest phases of the chemical enrichment of the Universe with the scatter observed in extremely-metal poor halo field stars[7] (with [Fe/H] < −3) and in the stars of NGC 6522 (with [Fe/H] = −1; see Supplementary Information for a discussion on other abundance ratios). Two important conclusions can be drawn. First, the observed scatter in [Y/Ba] in the early Universe (both in the Bulge and in the very-metal-poor halo stars) is compatible with the expected scatter from 'spinstar' models, apart from very few objects with [Y/Ba] ratios below −1, in the case of the halo field stars. Second, whereas 'spinstars' can produce a large scatter in the [Y/Ba] ratios, covering a range of −1 < [Y/Ba] < +2, AGB models[14] at [Fe/H] = −1 (compatible with the observed high [Y/Fe] and [Y/Ba] ratios in Table 1) cover a smaller [Y/Ba] range (−1 < [Y/Ba] < 0.5), and cannot account for the large [Y/Ba] ratios of stars number 3, 7 and 8 in Fig. 1 (see Supplementary Information for more details). In

addition, if we take into account the results shown in Fig. 1 together with the other element ratios presented in the Supplementary Information, a combination of the s-process component from 'spinstars' and the explosive r-process component may explain the heavy elements in all the stars observed in NGC 6522. Predictions from AGB models may explain the abundances in five out of eight stars (not reproducing those with the highest [Y/Ba]). However, the AGB mass-transfer scenario might have difficulty in explaining low [C/Fe] ratios. Hence it is of primary importance to obtain precise [C/Fe] ratios for the NGC 6522 stars, not just upper limits, in order to distinguish between 'spinstars' and AGB mass-transfer scenarios.

A possible observational test to confirm our scenario would be to look for the scatter in the [Pb/Fe] ratios in the early Universe. Low-metallicity AGB mass-transfer models[14,15] have quite a robust prediction for the minimum expected [Pb/Eu] ratios, but 'spinstar' models predict a large scatter in the [Pb/Eu] ratios. Therefore, if Pb could also be measured for stars with known abundances of Sr, Y, Zr, Ba and La, it would be possible to distinguish between the two scenarios. Unfortunately, the useful Pb lines are in the ultraviolet part of the spectrum and thus are highly extinct for bulge stars.

Other processes besides fast rotation have been invoked in the literature to explain the large overabundances with respect to Fe of the light s-process elements (Y, Sr and Zr) in the very-metal-poor Universe[16–18], but it is unclear if they could also account for the abundances observed in NGC 6522. Here we propose that 'spinstars' offer another plausible explanation for this signature, both in the Bulge and in extremely metal-poor halo stars.

The impact of having had an early generation of 'spinstars' in the Universe is manifold. They may have contributed to the primary
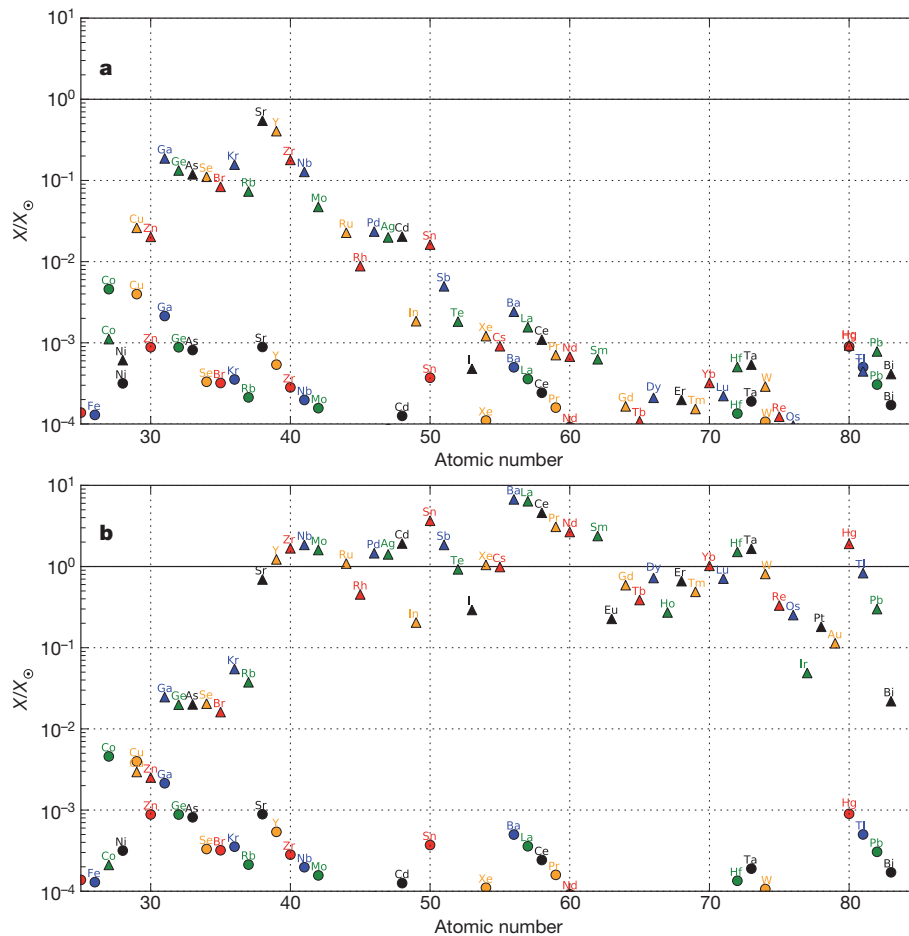
Figure 2 | Impact of rotation on s-process element production in very-metal-poor massive stars. The models show the production of s-process elements (where $X/X_\odot$ is the mass fraction of atoms normalized to solar) predicted in a $40M_\odot$ star with [Fe/H] = −3.8, which does not rotate (circles), and with an initial rotational velocity of 500 km s^{-1} (triangles). These models are shown at two different phases of He-burning: **a**, at the beginning of the s-process (about 127,000 years before He exhaustion); and **b**, at the end of the central He-burning phase. Rotation boosts production of the s-process elements by four orders of magnitude. $X_\odot$, solar abundance of given element.

nitrogen production in the early Universe[19,20]. The fast spins of the stars could have led to more mass loss than expected at these very low metallicities and thus could have prevented the first stars from dying as pair-instability supernovae[11], which would explain why halo stars do not bear the chemical signature of pair-instability supernovae. In addition, even if the more-massive 'spinstars' were to collapse into black holes without a supernova explosion, they would have been able to contribute to the chemical enrichment of the interstellar medium in the very early Universe via stellar mass-loss triggered by rotation. This also has implications for the fate of the stars, leading possibly to more gamma-ray bursts than previously thought, and more generally to magneto-hydrodynamic explosions[11]. Finally, 'spinstars' have longer lifetimes[11] and possibly higher surface temperatures and luminosities than non-rotating stars. It will therefore be worthwhile to study the impact of rotation on the ionizing power of the first stars. The fast rotation of the first stars is now also supported by the latest hydrodynamic simulations of the formation of the first stars[21].

1. Bromm, V., Yoshida, N., Hernquist, L. & McKee, C. The formation of the first stars and galaxies. *Nature* **459**, 49–54 (2009).
2. Truran, J. W. A new interpretation of the heavy element abundances in metal-deficient stars. *Astron. Astrophys.* **97**, 391–393 (1981).
3. Cowan, J. J. & Sneden, C. Heavy element synthesis in the oldest stars and the early Universe. *Nature* **440**, 1151–1156 (2006).
4. Barbuy, B. *et al.* VLT-FLAMES analysis of eight giants in the bulge metal-poor globular cluster NGC 6522: oldest cluster in the Galaxy? *Astron. Astrophys.* **507**, 405–415 (2009).
5. Sneden, C., Cowan, J. J. & Gallino, R. Neutron-capture elements in the early Galaxy. *Annu. Rev. Astron. Astrophys.* **46**, 241–288 (2008).
6. Pignatari, M. *et al.* The s-process in massive stars at low metallicity: the effect of primary ^{14}N from fast rotating stars. *Astrophys. J.* **687**, L95–L98 (2008).
7. Frebel, A. Stellar archaeology: exploring the Universe with metal-poor stars. *Astron. Nachr.* **331**, 474–488 (2010).
8. Beers, T. C. & Christlieb, N. The discovery and analysis of very metal-poor stars in the Galaxy. *Annu. Rev. Astron. Astrophys.* **43**, 531–580 (2005).
9. Martayan, C. *et al.* Be stars and binaries in the field of the SMC open cluster NGC 330 with VLT-FLAMES. *Astron. Astrophys.* **472**, 577–586 (2007).
10. Hirschi, R. Very low-metallicity massive stars: pre-SN evolution and primary nitrogen production. *Astron. Astrophys.* **461**, 571–583 (2007).
11. Ekström, S., Meynet, G., Chiappini, C., Hirschi, R. & Maeder, A. Effects of rotation on the evolution of primordial stars. *Astron. Astrophys.* **489**, 685–698 (2008).
12. Meynet, G., Ekström, S. & Maeder, A. The early star generations: the dominant effect of rotation on CNO yields. *Astron. Astrophys.* **447**, 623–639 (2006).
13. Frischknecht, U., Hirschi, R., Meynet, G., Ekstroem, S., Georgy, C., Rauscher, T., Winteler, C. & Thielemann, F.-K. Constraints on rotational mixing from surface evolution of light elements in massive stars. *Astron. Astrophys.* **522**, A39 (2010).
14. Bisterzo, S. *et al.* s-process in low metallicity stars – I. Theoretical predictions. *Mon. Not. R. Astron. Soc.* **404**, 1529–1544 (2010).
15. Karakas, A. I. Updated stellar yields from asymptotic giant branch models. *Mon. Not. R. Astron. Soc.* **403**, 1413–1425 (2009).
16. Qian, Y.-Z. & Wasserburg, G. J. Abundances of Sr, Y and Zr in metal-poor stars and implications for chemical evolution in the early Galaxy. *Astrophys. J.* **687**, 272–286 (2008).
17. Farouqi, K. *et al.* Nucleosynthesis modes in the high-entropy wind of type II supernovae: comparison of calculations with halo-stars observations. *Astrophys. J.* **694**, L49–L53 (2009).
18. Travaglio, C. *et al.* Galactic evolution of Sr, Y, and Zr: a multiplicity of nucleosynthetic processes. *Astrophys. J.* **601**, 864–884 (2004).
19. Chiappini, C. *et al.* A strong case for fast stellar rotation at very low metallicities. *Astron. Astrophys.* **449**, L27–L30 (2006).

20. Chiappini, C. *et al.* A new imprint of fast rotators: low $^{12}C/^{13}C$ ratios in extremely metal-poor halo stars. *Astron. Astrophys.* **479,** L9–L12 (2008).
21. Stacy, A., Bromm, V. & Loeb, A. Rotation seed of the first stars. *Mon. Not. R. Astron. Soc.* (in the press).

LETTER

Superconductor–insulator transition in La$_{2-x}$Sr$_x$CuO$_4$ at the pair quantum resistance

A. T. Bollinger[1], G. Dubuis[1,2], J. Yoon[1], D. Pavuna[2], J. Misewich[1] & I. Božović[1]

High-temperature superconductivity in copper oxides arises when a parent insulator compound is doped beyond some critical concentration; what exactly happens at this superconductor–insulator transition is a key open question[1]. The cleanest approach is to tune the carrier density using the electric field effect[2–7]; for example, it was learned in this way[5] that weak electron localization transforms superconducting SrTiO$_3$ into a Fermi-glass insulator. But in the copper oxides this has been a long-standing technical challenge[3], because perfect ultrathin films and huge local fields ($>10^9$ V m^{-1}) are needed. Recently, such fields have been obtained using electrolytes or ionic liquids in the electric double-layer transistor configuration[8–10]. Here we report synthesis of epitaxial films of La$_{2-x}$Sr$_x$CuO$_4$ that are one unit cell thick, and fabrication of double-layer transistors. Very large fields and induced changes in surface carrier density enable shifts in the critical temperature by up to 30 K. Hundreds of resistance versus temperature and carrier density curves were recorded and shown to collapse onto a single function, as predicted for a two-dimensional superconductor–insulator transition[11–14]. The observed critical resistance is precisely the quantum resistance for pairs, $R_Q = h/(2e)^2 = 6.45$ kΩ, suggestive of a phase transition driven by quantum phase fluctuations, and Cooper pair (de)localization.

We used atomic-layer molecular beam epitaxy to synthesize a number of ultrathin films of La$_{2-x}$Sr$_x$CuO$_4$ (LSCO) that were 1, 1.5 or 2 unit cells thick with $x = 0.06$ to 0.20, on top of some insulating La$_2$CuO$_4$ buffer layers grown on LaSrAlO$_4$ substrates[15–17] (see Supplementary section A). Devices of well-defined geometry were fabricated lithographically (Supplementary section B) for accurate measurements of resistivity and magnetic susceptibility. Various electrolytes (Supplementary section C) were explored to find the one affecting film quality the least. The total number of films we used was about 30, with several hundred devices, of which over a hundred were tested extensively. In some cases, hundreds of measurements were made on the same device—the process is generally reversible and non-destructive, unless some threshold voltage is exceeded.

Upon applying a gate voltage (V_G), large shifts are observed both in the normal state resistivity and in the critical temperature (T_c), see Fig. 1a. In this and other samples, in the top layer we see shifts in the induced carrier density x per Cu atom (see Supplementary section D) by up to ±0.04 and shifts in T_c (defined as the transition midpoint) by as much as 30 K—more than 75% of the full range from about 2 K (the low-T limit of our measurements) to the maximum ($T_c^{max} \approx 40$ K) that we have observed in thicker single-phase LSCO films at optimum doping. The field effect works in both directions: we can achieve electron accumulation or depletion, and increases or decreases in T_c, in the expected direction, depending on whether the sample was initially overdoped or underdoped. The diamagnetic response, measured by two-coil mutual inductance technique, also shows large shifts in T_c when the gate voltage is applied (see Fig. 1b). This demonstrates complete superconductivity: strong diamagnetic screening begins when the supercurrent closes a complete loop at or below $T_c(R = 0)$.

To address the issue of chemical interdiffusion, we have varied the film thickness and verified that it was not possible to reduce T_c in
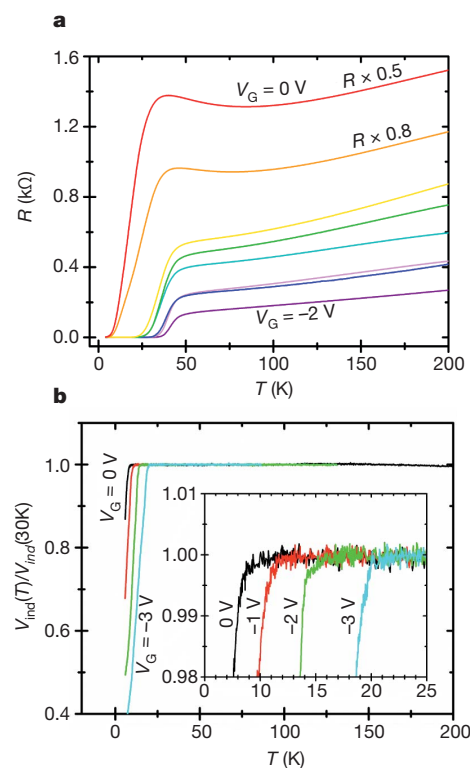

Figure 1 | Response of superconducting field effect devices to the gate voltage. Fabrication of a superconducting field-effect transistor poses two tough technical challenges[3]. First, LSCO has $n = 1.5\ 10^{21}$ cm^{-3}, much higher than a typical semiconductor, so doping a single CuO$_2$ layer to this optimum level requires huge local electric field ($>10^9$ V m^{-1}). Next, in LSCO the Thomas–Fermi screening length is much shorter[17] ($\lambda_{TF} = 6 \pm 2$ Å) than in semiconductors, so the second CuO$_2$ layer below the surface remains heavily underdoped; therefore one needs perfect ultrathin (~1 nm thick) HTS films. Here, we make use of two technical advances: atomic-layer molecular beam epitaxy[15–17] synthesis, and the electric double-layer transistor technique[8–10] inspired by electrolyte supercapacitors, in which electric fields of 10^{10} V m^{-1} are attained inside the Helmholtz double layers while the induced surface charge density reaches 10^{14}–10^{15} cm^{-2}. (For details of device design see Supplementary section B.) **a,** Temperature dependence of resistance for a device using a Pt top gate and polymer electrolyte (NaF dissolved in polyethylene glycol). Initially the active layer is underdoped. At the temperature above the melting point of the dielectric, the hole density increases as the gate voltage is ramped from 0 V to −2 V. Charge is accumulated for a fixed time (typically 30 min) and then frozen in place by cooling the sample. This induces a shift in T_c of nearly 30 K. **b,** Temperature dependence of mutual inductance for another device using an Au top gate and the ionic liquid electrolyte N,N-diethyl-N-methyl-N-(2-methoxyethyl)ammonium bis(trifluoromethylsulphonyl)imide (DEME-TFSI). The film is placed between two coils; when superconducting current loops are closed, the diamagnetic response reduces the mutual inductance. The gate voltage is applied from 0 V to −3 V in 1 V steps. The inset is a magnified view near the transition, demonstrating the increase in T_c.

[1]Brookhaven National Laboratory, Upton, New York 11973-5000, USA. [2]Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland.

optimally doped films thicker than a few unit cells; in this case, the deeper layers stay unaffected and shunt the few top layers that are modified by the field. Also, we explored various electrolytes (Supplementary section C) and found essentially no difference, despite the active ions being very different. Finally, low-energy electron microscopy and diffraction showed no evidence of any of the surface changes that one would expect in the case of ion insertion.

A large number of $R(T, V_G)$ curves were recorded, scanning V_G in small steps. In some initially underdoped samples, the data span the superconductor–insulator transition in a broad range, with many curves on both sides; an example is shown in Fig. 2a. Such data enable a meaningful scaling analysis that can teach us about the nature of the quantum phase transition (QPT) and the quantum critical behaviour in underdoped films. The theory of continuous phase transitions[11] states that in the critical region physical observables obey scaling laws such that their correlation functions depend on space and time coordinates only in combinations r/ξ and t/τ, where the scales ξ (correlation length) and τ (correlation time) are determined by microscopic parameters of the Hamiltonian and diverge at the transition as $\xi \propto |x - x_c|^{-v}$ and $\tau \propto \xi^z \propto |x - x_c|^{-zv}$. The actual scaling functions, together with the exponents v and z, depend only on the large-scale physics, the universality class being determined by such general properties as the space dimensionality, presence of disorder, and the symmetry of the order parameter manifold. Competition between energy and entropy can drive a classical phase transition at finite temperature $T = T_c$. QPTs, on the other hand, occur at $T = 0$, triggered by a change of some external parameter (pressure, chemical composition, etc.). A quantum critical point separates ground states with different symmetry. The scaling of resistivity near the superconductor–insulator transition in (quasi-) two-dimensional systems is of particular interest because two-dimensionality is special (it is the lower critical dimension) for both superconductivity and localization, and because in two dimensions, the length scale drops out for resistance so that its value at the critical point should be universal[11–14]. For two-dimensional QPT, the appropriate finite-size scaling is $R_\square(x, T) = R_c F(|x - x_c|T^{-1/zv})$, where $R_\square$ is the resistance per square, R_c is the limiting ($x \to x_c$ and $T \to 0$) resistance, and $F(u)$ is some universal function of u (≥ 0) such that $F(u) \to 1$ when $u \to 0$. The value of R_c, which can be obtained by scaling from data taken at finite T, together with v and z encodes the essential physics of the transition.

In Fig. 2b, we re-plot the same $R_\square$ data as a function of a measured quantity, $x = 0.33\,\mathrm{k\Omega}/R_\square(T = 180\,\mathrm{K})$, approximately equal to the number of mobile holes per one formula unit (Supplementary section D). (The precise value of carrier density affects none of our conclusions.) We note that this is the same $R_\square(T, x)$ data matrix as in Fig. 2a, but we have inverted it here, that is, we plot $R_\square(x)$ for various values of T. All curves are seen to cross at a single point x_c, which corresponds to the separatrix, that is, the $R(T)$ curve that is essentially constant in the temperature interval under study, and which separates insulating from superconducting behaviour. This can be seen directly in the inset to Fig. 2a. From Fig. 2b, we can read the critical values of x and $R_\square$ to be $x_c \approx 0.06$ and $R_c = 6.45 \pm 0.10\,\mathrm{k\Omega}$. In Fig. 2c, we have scaled the abscissa as $u = |x - x_c|T^{-1/zv}$. As the result of this scaling, hundreds of curves have collapsed to one, showing at $T = 0$ a bifurcation, or unstable fixed point. The collapse is excellent up to about 10 K for $zv = 1.5 \pm 0.1$, and the exponent is identical on both sides of the superconductor–insulator transition. An alternative method of determining zv gave the same result (Supplementary section F). This is consistent with a continuous two-dimensional superconductor–insulator QPT, with an extensive region of quantum critical behaviour. The good agreement with a two-dimensional theory[11–14] suggests that the active part of the film is extremely thin, in contrast to what one would expect in the case of massive interdiffusion or intercalation, for which the active part of the film is certainly deeper and probably fractal-like.

The nature of the two-dimensional superconductor–insulator QPT has been intensely debated[11–14,18–24]. Roughly, the models can be grouped into two classes, fermionic (appropriate for weak coupling) and bosonic (appropriate for strong coupling), plus some that incorporate both aspects. The more conventional, fermionic picture envisions the superconductor–insulator transition driven by amplitude fluctuations; on the insulating side, the Cooper pairs are broken and the electrons are localized owing to a random potential (Anderson localization), interactions (Mott insulator), or a combination of both. This localized state—a Fermi glass—is gauge invariant, while in the superconducting state the global and the local gauge symmetry are broken. The alternative, bosonic picture[12] assumes that in the superconducting phase mobile Cooper pairs form the superfluid condensate, whereas vortices are localized (forming bound vortex–antivortex pairs). On the insulating side, the vortex–antivortex pairs are broken (at low T, by quantum fluctuations) and mobile, thus causing dissipation and finite
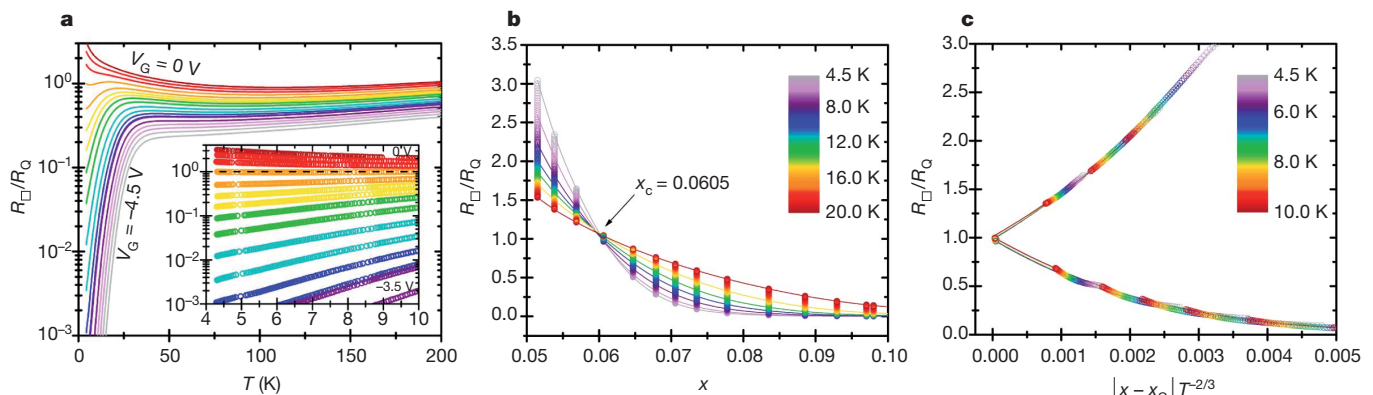

Figure 2 | Superconductor–insulator transition driven by electric field.
a, Temperature dependence of normalized resistance $r = R_\square(x, T)/R_Q$ of an initially heavily underdoped and insulating film (see Supplementary Fig. 12 for linear scale). The device (Supplementary section B) employs a coplanar Au gate and DEME-TFSI ionic liquid. The carrier density, fixed for each curve, is tuned by varying the gate voltage from 0 V to -4.5 V in 0.25 V steps; an insulating film becomes superconducting via a QPT. The inset highlights a separatrix independent of temperature below 10 K. The open circles are the actual raw data points; the black dashed line is $R_\square(x_c, T) = R_Q = 6.45\,\mathrm{k\Omega}$. **b**, The inverse representation of the same data, that is, the $r_T(x)$ dependence at fixed temperatures below 20 K. Each vertical array of (about 100) data points corresponds to one fixed carrier density, that is, to one $r_x(T)$ curve in Fig. 2a.

The colours refer to the temperature, and the continuous lines are interpolated for selected temperatures (4.5, 6.0, 8.0, 10.0, 12.0, 15.0 and 20.0 K). The crossing point defines the critical carrier concentration $x_c = 0.06 \pm 0.01$, and the critical resistance $R_c = 6.45 \pm 0.10\,\mathrm{k\Omega}$. **c**, Scaling of the same data with respect to a single variable $u = |x - x_c|T^{-1/zv}$, with $zv = 1.5$. This figure is derived by folding panel **b** at x_c and scaling the abscissa of each $r_T(|x - x_c|)$ curve by $T^{-2/3}$. For $4.3\,\mathrm{K} < T < 10\,\mathrm{K}$, the discrete groups of points of Fig. 2b collapse accurately onto a two-valued function, with one branch corresponding to x larger and the other to x smaller than x_c. The critical exponents are identical on both sides of the superconductor–insulator transition. The raw data points cover the interpolation lines almost completely, except close to the origin.

resistance, whereas the Cooper pairs are localized and rendered immobile by disorder, forming a Bose glass. The fermionic degrees of freedom are ignored, at least near the QPT. The scaling itself is not sufficient to differentiate between different models; the truth must lie in the details, such as the values of R_c, v and z. Within the bosonic scenario, R_c should be a universal constant; if the duality between vortices and Cooper pairs was perfect, R_c should be equal to the quantum resistance for pairs, $R_Q \equiv h/(2e)^2 = 6.45\,k\Omega$.

Our experimental results are in line with the predictions of the bosonic theory in its simplest (self-dual) form. In particular, $R_c = R_Q \equiv h/(2e)^2$ seems very telling; the simplest explanation is that pairs exist on both sides of the superconductor–insulator QPT, that is, that the transition is driven by quantum phase fluctuations. This physical picture requires the existence of a Bose glass of localized pairs, and of freely moving vortices, on the insulating side of the QPT. This is consistent with some peculiar experimental observations: mobile vortices were observed by torque magnetometry in LSCO down to very low doping ($x = 0.03$), well within the insulating phase[25], while scanning tunnelling microscopy showed signs of localized pairs in insulating $La_{1.88}Ba_{0.12}CuO_4$ samples[26]. Terahertz spectroscopy indicates that in the 'vertical' (that is, T-driven) transitions high-temperature superconductivity (HTS) is destroyed by thermal phase fluctuations, everywhere in the phase diagram[27]. The present work complements this picture, by showing that in the 'horizontal' (x-driven) transitions at low T, superconductivity is destroyed by quantum phase fluctuations. We note that the superfluid carrier density at $T = 0\,K$, n_s, which is a measure of the phase stiffness, is two orders of magnitude smaller in copper oxides than in conventional superconductors, hence one should expect the former to be much more susceptible to phase fluctuations[28]. This should be generic to HTS copper oxides; indeed, microwave[29] and mutual inductance[30] measurements on strongly underdoped $YBa_2Cu_3O_{6+y}$ showed that T_c scales as $\sqrt{n_s}$, also pointing to quantum critical behaviour. The difference is that in $YBa_2Cu_3O_{6+y}$ fluctuations appear to be correlated over many unit cells in the c-axis direction[29,30].

Even compared to low-temperature superconductors, we find some similarities; for example, Fig. 2c (or Supplementary Fig. 14, which shows the same data on the log scale) bears a resemblance to the scaled $R(T)$ curves of thin films of conventional superconductors[2,5–8,13,14,19,20,23] such as $SrTiO_3$ or α-MoGe. This is expected, because all these systems show two-dimensional superconductor–insulator QPT. But the detailed physics, microscopic as well as large-scale, in the HTS films studied here is essentially different. Superconductivity in $SrTiO_3$ and α-MoGe is mediated by weak electron–phonon coupling, the microscopic mechanism of which is quantitatively accounted for by the Bardeen–Cooper–Schrieffer theory. In contrast, the mechanism of HTS in copper oxides is not known, although it is clearly different from the Bardeen–Cooper–Schrieffer electron–phonon mechanism. Moreover, the superconductor–insulator transition in amorphous α-MoGe is driven[19,20] by quantum percolation with the critical exponents product $zv = 4/3$, whereas in $SrTiO_3$ we find $zv = 2/3$ and the superconductor–insulator transition to (Fermi-glass) insulating behaviour driven by weak localization[5]. The critical resistance is universal neither in α-MoGe nor in $SrTiO_3$, and can be ten times smaller than R_Q. The superconductor–insulator transition in LSCO is in a different universality class, with $R_c = R_Q$ and $zv = 3/2$. The films are single-crystal, neither amorphous nor granular; the transition seems to be $(2e)$-to-$(2e)$, that is, the Bose-glass insulating state stems from localized pairs. We hope that this insight brings us a step closer to understanding HTS.

1. Lee, P. A., Nagaosa, N. & Wen, X. G. Doping a Mott insulator: physics of high-temperature superconductivity. *Rev. Mod. Phys.* **78**, 17–85 (2006).
2. Parendo, K. A. *et al.* Electrostatic tuning of the superconductor-insulator transition in two dimensions. *Phys. Rev. Lett.* **94**, 197004 (2005).
3. Ahn, C. H. *et al.* Electrostatic modification of novel materials. *Rev. Mod. Phys.* **78**, 1185–1212 (2006).
4. Matthey, D., Reyren, N., Triscone, J.-M. & Schneider, T. Electric-field-effect modulation of the transition temperature, mobile carrier density, and in-plane penetration depth of NdBa₂Cu₃O₇₋δ thin films. *Phys. Rev. Lett.* **98**, 057002 (2007).
5. Caviglia, A. D. *et al.* Electric field control of the LaAlO₃/SrTiO₃ interface ground state. *Nature* **456**, 624–627 (2008).
6. Bell, C. *et al.* Dominant mobility modulation by the electric field effect at the LaAlO₃/SrTiO₃ Interface. *Phys. Rev. Lett.* **103**, 226802 (2009).
7. Biscaras, J. *et al.* Two-dimensional superconductivity at a Mott insulator/band insulator interface LaTiO₃/SrTiO₃. *Nature Commun.* **1**, 1–5 (2010).
8. Ueno, K. *et al.* Electric-field-induced superconductivity in an insulator. *Nature Mater.* **7**, 855–858 (2008).
9. Ye, J. T. *et al.* Liquid-gated interface superconductivity on an atomically flat film. *Nature Mater.* **9**, 125–128 (2010).
10. Dhoot, A. S. *et al.* Increased T_c in electrolyte-gated cuprates. *Adv. Mater.* **22**, 2529–2533 (2010).
11. Sondhi, S. L., Grivin, S. M., Carini, J. P. & Shahar, D. Continuous quantum phase transitions. *Rev. Mod. Phys.* **69**, 315–333 (1997).
12. Fisher, M. P. A., Grinstein, G. & Girvin, S. M. Presence of quantum diffusion in two dimensions: universal resistance at the superconductor-insulator transition. *Phys. Rev. Lett.* **64**, 587–590 (1990).
13. Goldman, A. M. & Marković, N. Superconductor-insulator transitions in the two-dimensional limit. *Phys. Today* **51**, 39–44 (1998).
14. Gantmakher, V. F. & Dolgopolov, V. T. Superconductor-insulator quantum phase transition. *Phys. Usp.* **53**, 1–49 (2010).
15. Gozar, A. *et al.* High-temperature interface superconductivity between metallic and insulating copper oxides. *Nature* **455**, 782–785 (2008).
16. Logvenov, G., Gozar, A. & Bozovic, I. High-temperature superconductivity in a single copper-oxygen plane. *Science* **326**, 699–702 (2009).
17. Smadici, S. *et al.* Superconducting transition at 38 K in insulating-overdoped La₂CuO₄-La₁.₆₄Sr₀.₃₆CuO₄ superlattices: evidence for interface electronic redistribution from resonant soft X-ray scattering. *Phys. Rev. Lett.* **102**, 107004 (2009).
18. Wang, T. *et al.* Onset of high-temperature superconductivity in the two-dimensional limit. *Phys. Rev. B* **43**, 8623–8626 (1991).
19. Yazdani, A. & Kapitulnik, A. Superconducting-insulating transition in two-dimensional α-MoGe thin films. *Phys. Rev. Lett.* **74**, 3037–3040 (1995).
20. Mason, N. & Kapitulnik, A. Dissipation effects on the superconductor-insulator transition in 2D superconductors. *Phys. Rev. Lett.* **82**, 5341–5344 (1999).
21. Herbut, I. F. Critical exponents at the superconductor-insulator transition in dirty-boson systems. *Phys. Rev. B* **61**, 14723–14726 (2000).
22. Phillips, P. & Dalidovich, D. The elusive Bose metal. *Science* **302**, 243–247 (2003).
23. Steiner, M. A., Breznay, N. P. & Kapitulnik, A. Approach to a superconductor-to-Bose-insulator transition in disordered films. *Phys. Rev. B* **77**, 212501 (2008).
24. Tešanović, Z. d-wave duality and its reflections in high-temperature superconductors. *Nature Phys.* **4**, 408–414 (2008).
25. Li, L., Checkelsky, J. G., Komiya, S., Ando, Y. & Ong, N. P. Low-temperature vortex liquid in La₂₋ₓSrₓCuO₄. *Nature Phys.* **3**, 311–314 (2007).
26. Valla, T., Fedorov, A. V., Lee, J., Davis, J. C. & Gu, G. D. The ground state of the pseudogap in cuprate superconductors. *Science* **314**, 1914–1916 (2006).
27. Bilbro, L. S. *et al.* Temporal correlations of superconductivity above the transition temperature in La₂₋ₓSrₓCuO₄ probed by terahertz spectroscopy. *Nature Phys.* **7**, 298–302 (2011).
28. Emery, V. J. & Kivelson, S. A. Importance of phase fluctuations in superconductors with small superfluid density. *Nature* **374**, 434–437 (1995).
29. Broun, D. M. *et al.* Superfluid density in a highly underdoped YBa₂Cu₃O₆₊y superconductor. *Phys. Rev. Lett.* **99**, 237003 (2007).
30. Hetel, I., Lemberger, T. R. & Randeria, M. Quantum critical behaviour in the superfluid density of strongly underdoped ultrathin copper oxide films. *Nature Phys.* **3**, 700–702 (2007).

Author Contributions I.B. and A.T.B. conceived and designed the experiments. I.B. synthesized the films. A.T.B. fabricated the devices. A.T.B., G.D. and J.Y. carried out the transport measurements. All authors analysed the results and contributed to writing the manuscript.

LETTER

Continuing Colorado plateau uplift by delamination–style convective lithospheric downwelling

A. Levander[1], B. Schmandt[2], M. S. Miller[3], K. Liu[1], K. E. Karlstrom[4], R. S. Crow[4], C.-T. A. Lee[1] & E. D. Humphreys[2]

The Colorado plateau is a large, tectonically intact, physiographic province in the southwestern North American Cordillera that stands at ~1,800–2,000 m elevation and has long been thought to be in isostatic equilibrium[1]. The origin of these high elevations is unclear because unlike the surrounding provinces, which have undergone significant Cretaceous–Palaeogene compressional deformation followed by Neogene extensional deformation, the Colorado plateau is largely internally undeformed. Here we combine new seismic tomography[2] and receiver function images to resolve a vertical high-seismic-velocity anomaly beneath the west-central plateau that extends more than 200 km in depth. The upper surface of this anomaly is seismically defined by a dipping interface extending from the lower crust to depths of 70–90 km. The base of the continental crust above the anomaly has a similar shape, with an elevated Moho. We interpret these seismic structures as a continuing regional, delamination-style foundering of lower crust and continental lithosphere. This implies that Pliocene (2.6–5.3 Myr ago) uplift of the plateau and the magmatism on its margins are intimately tied to continuing deep lithospheric processes. Petrologic and geochemical observations indicate that late Cretaceous–Palaeogene (~90–40 Myr ago) low-angle subduction hydrated and probably weakened much of the Proterozoic tectospheric mantle[3–5] beneath the Colorado plateau. We suggest that mid-Cenozoic (~35–25 Myr ago) to Recent magmatic infiltration subsequently imparted negative compositional buoyancy to the base and sides of the Colorado plateau upper mantle, triggering downwelling. The patterns of magmatic activity suggest that previous such events have progressively removed the Colorado plateau lithosphere inward from its margins[6], and have driven uplift. Using Grand Canyon incision rates[7,8] and Pliocene basaltic volcanism patterns, we suggest that this particular event has been active over the past ~6 Myr.

Unlike the basement-involved Laramide uplifts of the Rocky Mountains (~2,200 m elevation) and the 100% extension of the Basin and Range (at 1,000–1,600 m elevation) since the mid-Cenozoic, the Colorado plateau (Fig. 1) has experienced only mild internal deformation (~1% shortening), manifested as long-wavelength monoclines. Xenolith data indicate that central Colorado plateau mantle consisted of Proterozoic (~1.7–2.0 Gyr old) North American tectospheric mantle extending to 120 km depth as recently as 20–40 Myr ago[4,9]. In the Late Miocene and Pliocene, the western and southern peripheries of the plateau were invaded by basaltic magmas[6] whose Nd and Sr isotopes indicate an increasingly undepleted mantle source[7,10].

The timing of both Colorado plateau uplift and formation of the Grand Canyon remain unresolved, with various lines of evidence suggesting episodes of uplift in the Laramide[11,12], mid-Cenozoic[3,13] and late Cenozoic[8,14,15]. The cumulative uplift history probably reflects lithospheric preconditioning from hydration of the North American lithosphere during Farallon plate flat-slab subduction that weakened the overriding plate, possibly removing any North American lithosphere deeper than 120–150 km, and cooling the remainder from

below. Models for late Cenozoic steady magmatic invasion of the northwestern and southern margins towards the core of the Colorado plateau and the related plateau uplift are attributed to
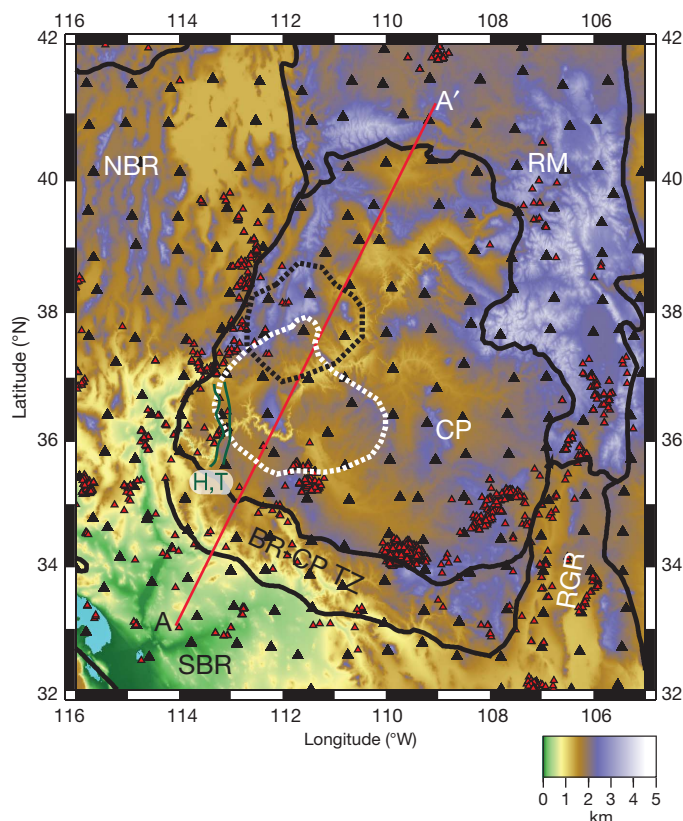

Figure 1 | Elevation map of the Colorado plateau. The map shows topography, in km, of the Colorado plateau and surrounding tectonic/physiographic provinces, province boundaries (black lines), and the locations of approximately 220 seismograph stations. CP, Colorado plateau; NBR, northern Basin and Range; SBR, southern Basin and Range; RGR, Rio Grande rift; RM, Rocky Mountains; BR-CP TZ, Basin and Range-Colorado Plateau transition zone. H, Hurricane fault; T, Toroweap fault. Black triangles, USArray Transportable Array and other broadband seismograph stations used in the seismic analysis; red triangles, volcanic rocks less than 10 Myr old. Profile AA′ is the location of the cross-sections in Fig. 3. White dashed line, region of lower-crustal delamination and crustal thinning, with the western edge coinciding with active normal faults (blue lines) that mark ~100 m Myr^{-1} differential uplift of the Colorado plateau relative to BR. Young volcanics ring the delamination region. Black dashed line, outline of the downwelling body at 200 km depth estimated from the body wave tomography (Figs 3 and 4). The Grand Canyon is centred on the delamination region, suggesting that the downwelling influenced re-routing of Rocky Mountain and Colorado plateau headwaters into a through-going Colorado River that carved the Grand Canyon in the last 5–6 Myr (ref. 8).

[1]Earth Science Department, Rice University, Houston, Texas 77005-1892, USA. [2]Department of Geological Sciences, University of Oregon, Eugene, Oregon 97403, USA. [3]Department of Earth Sciences, University of Southern California, Los Angeles, California 90089-0740, USA. [4]Department of Earth and Planetary Sciences, University of New Mexico, Albuquerque, New Mexico 87131, USA.

conductive heating and thermal expansion[6] and small-scale convective removal of Colorado plateau lithospheric mantle[8].

Here we review seismic evidence that suggests how asthenosphere-derived melts are invading the Colorado plateau, and present evidence that suggests how these melts destabilize continental lithosphere and drive uplift. Previous active and passive seismic probes image shallow low-velocity zones (LVZ) under the southwestern Colorado plateau at 40–50 km depth (P-wave velocity $v_P \approx 7.5$ km s^{-1}) and under its eastern margin at 50–75 km depth ($v_P \approx 7.6$–7.7 km s^{-1}), with both LVZs lying beneath thin lithospheric lids with $v_P \approx 7.9$ km s^{-1} (refs 16, 17). Recent USArray Transportable Array P_n tomography measured upper-mantle velocities of ~8.0 km s^{-1} in the centre of the Colorado plateau, with 7.9 km s^{-1} and lower velocities along its western and southern peripheries[18].

We use PdS and SdP receiver functions to determine discontinuity structure beneath the Colorado plateau and its surroundings, and finite-frequency P and S body wave and Rayleigh wave tomography to construct three-dimensional models of v_P, v_S and v_P/v_S under the plateau[19]. The data for all these studies were recorded by the Transportable Array, other portable array experiments, and regional networks (Fig 1). The velocity model used for depth-positioning the receiver functions was constructed by smoothly blending Crust2.0

(ref. 20) with the upper-mantle P_n (ref. 18) and deeper-mantle teleseismic tomography[19] models.

Finite-frequency body-wave tomography finds low v_P, very low v_S and high v_P/v_S at ~60–200 km beneath the periphery of the Colorado plateau, forming a ring around its eastern, southern and northwestern edges, surrounding a higher-velocity, lower-v_P/v_S core (Fig. 2). Rayleigh wave tomography images a similar pattern of low shear velocities under the edges of the Colorado plateau, but with somewhat greater vertical resolution. Figure 2 compares the v_P and v_P/v_S anomalies (perturbations) and Rayleigh wave shear velocities at 80 km depth, and the depth of the lithosphere–asthenosphere boundary (LAB) from receiver functions and the Rayleigh wave v_S model. We interpret the low-velocity ring as low-melt-fraction asthenosphere surrounding the Colorado plateau lithospheric core.

Along profile AA′ (Fig. 1), the receiver function discontinuity structures and v_P and v_S tomography images show a set of structures suggestive of lower-crust and upper-mantle delamination (Fig. 3). The PdS receiver functions show a clearly defined Moho at 35–40 km and a bright sub-Moho event dipping ~10° northeast from ~50 to 90 km depth, between 35.5° and 38.0° N. Above the dipping event, a weaker event that we interpret as a 'new' Moho is elevated by approximately the thickness of the dipping event (10 km) over the same ~280 km
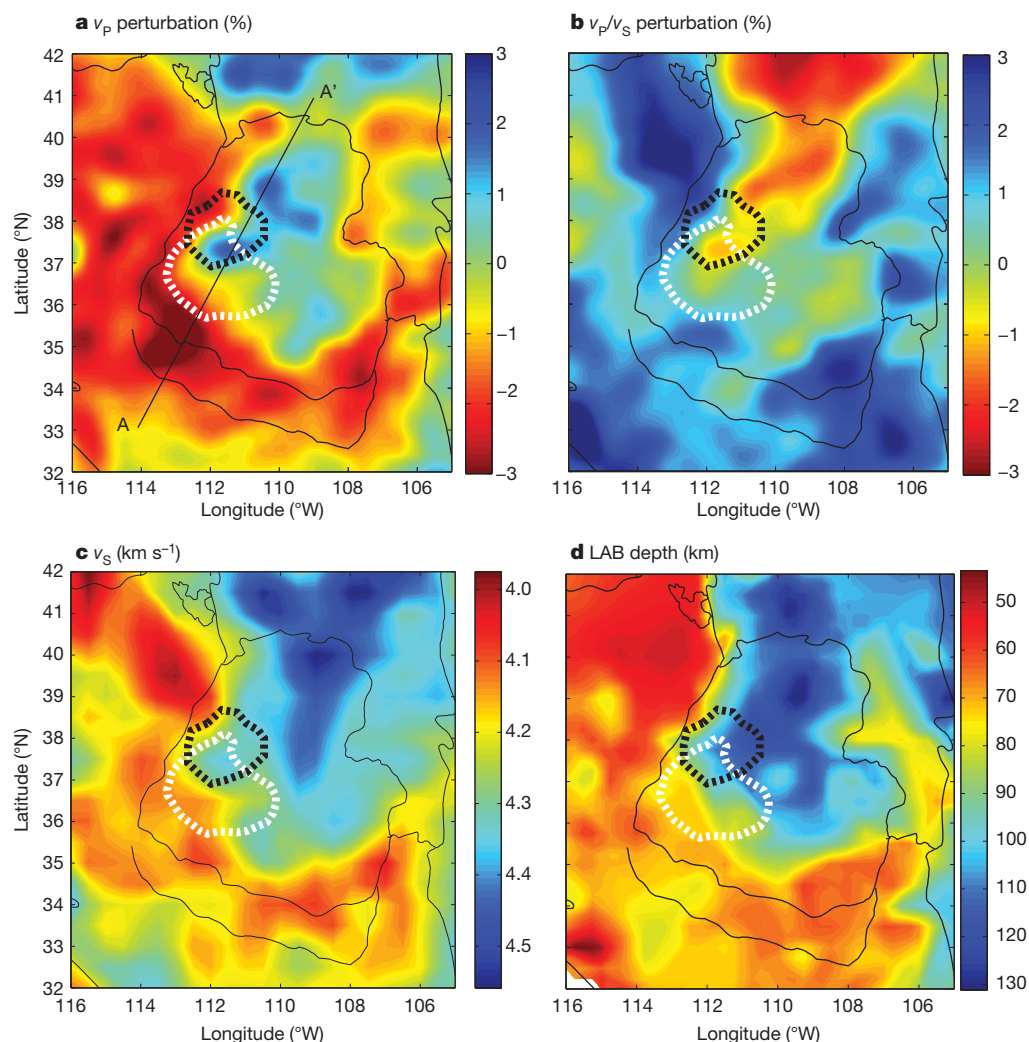

Figure 2 | Seismic velocity measurements at 80 km depth, and LAB depth, as measured with different seismic probes. a, b, Perturbations of v_P (**a**) and v_P/v_S (**b**) at 80 km depth determined by teleseismic body wave tomography. **c**, Shear velocity at 80 km depth measured by Rayleigh wave tomography. **d**, LAB depth determined from the receiver functions and the maximum negative vertical gradients in v_S as determined from the Rayleigh wave inversions. The Colorado

plateau is partially encircled by low v_P and v_S, and high v_P/v_S, suggestive of low-volume partial melt. The LAB is at shallow depths in the regions of low v_P and v_S, and high v_P/v_S, and deepens rapidly to the north under the central Colorado plateau and the Rocky Mountains. The locations of the crustal delamination region and the outline of the downwelling are shown as in Fig. 1.
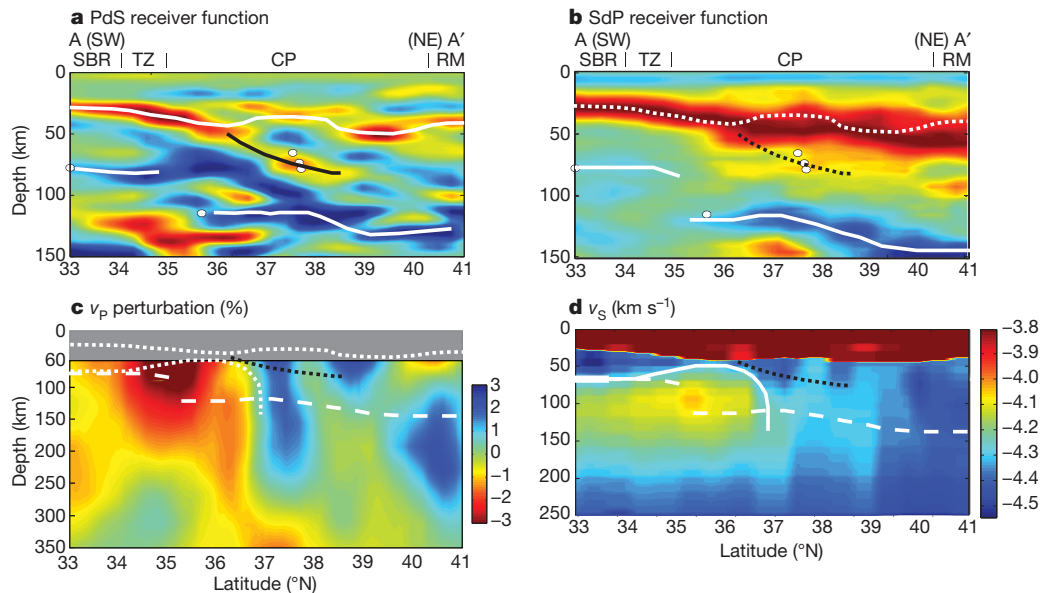
Figure 3 | Receiver function and seismic velocity cross-sections through the downwelling region. In all panels, the solid lines indicate surfaces chosen on that section, dashed lines are superimposed surfaces picked from the other sections. **a, b,** PdS (**a**) and SdP (**b**) receiver function sections along AA′ (Fig. 1). The PdS section shows a bifurcation in the Moho at ~36° N, with one branch shoaling to ~30 km and the other (black line) descending to 80–90 km. A negative polarity signal appears above the positive dipping event. At the same depths, the SdP receiver function shows a broadened positive lobe, the result of interference of two long period (5 s) pulses (Supplementary Information). The estimated melting depths of basalts along this profile are shown as white circles, and appear on either side of the deeper (black) surface.

The LAB is shown by the two deeper solid white lines in **a** and the two solid white lines in **b**. **c,** The v_P perturbation image shows a high-velocity downwelling that extends to >200 km depth, with necking at about 100 km depth. Note the vertical scale change. The downwelling is directly below the Moho bifurcation shown in **a**. **d**, v_S from Rayleigh wave tomography shows a pronounced upper-mantle LVZ in the southwestern part of the line to 112° W. Northeast of this, at ~100 km depth, v_S increases relative to the south. The LAB determined from the Rayleigh wave tomography is shown as a solid white line in **d** and as a short-dashed white line in **c**. The LAB from the SdP cross-section is shown as a long-dashed white line in **c** and **d**.

width. These two P to S conversions are apparent on a large number of stations (Supplementary Fig. 1a, b). The pattern suggests to us that the dipping structure once fitted in the Moho notch above it; this suggestion is supported by our examination of the dipping event and the shallow Moho in three dimensions (Supplementary Fig. 2), where we find approximately the same volume in the top of the dipping event as that beneath the elevated Moho and the surrounding Moho. We suggest that the lowermost crust and uppermost mantle are delaminating from the base of the Colorado plateau, as suggested previously[21], and as modelled in some two-dimensional geodynamic models[21,22].

Although this dipping structure is not as clear in the SdP receiver functions, the section shows complications beneath the Moho in the same places as the dipping event in the PdS receiver functions throughout the region outlined in Figs 1 and 2. We note that the resolution of SdP receiver functions is ~5 times coarser than that of PdS owing to differences in frequency content, and the blurred SdP image is expected from the lower frequency SdP receiver functions (Supplementary Fig. 3).

Directly beneath the dipping sub-Moho structure in the receiver function image, body wave tomography shows an elongate high-velocity body extending from 60 to greater than 200 km depth, with necking at about 100 km depth. The body wave tomography also shows high velocities and low v_P/v_S beneath the core of the Colorado plateau, typical of cratonic mantle lithosphere (Fig. 2). The Rayleigh wave v_S tomography model shows a strong LVZ ~100 km to the southwest of the drip region that changes to a pattern of intermediate and high velocities consistent with the receiver function sections. To the north, Rayleigh wave v_S becomes uniformly faster beneath the Colorado plateau core. Similar patterns appear in cross-sections taken at different azimuths through this area (Supplementary Fig. 4). Thus, using multiple seismic data analyses and techniques we identify this anomalous region (Fig. 1) as a currently occurring delamination-style convective downwelling, the causes of which we describe below.

The Moho and the deeper surfaces suggest that the currently active lithospheric foundering has propagated northeast to southwest, consistent with Grand Canyon uplift rates since 6 Myr ago inferred by its incision history, which indicates more rapid uplift in the eastern canyon (175–250 m Myr^{-1}) than in the western canyon (50–80 m Myr^{-1})[8]. Simple thermally based isostatic calculations on endmember models suggest that ~400–800 m of uplift would be possible if the entire Colorado plateau lithosphere were to be instantly replaced with asthenosphere, either with or without asthenospheric replacement of 10 km of lower crust (Supplementary Fig. 5). This estimated range is consistent with geologic evidence for uplift of the western Colorado plateau since 6 Myr ago. The amount of uplift is proportional to the ratio of the density of Colorado plateau lithosphere to that of asthenosphere. We describe below a means by which the density of the Colorado plateau lithosphere could have been increased.

Commonly suggested driving mechanisms for lithospheric foundering are: convective instability, with the lithosphere replaced by asthenosphere following orogenic thickening or some other perturbation of thermal boundary layers[21,23,24], development of eclogitic roots by the generation of mafic cumulates and restites during arc magmatism[25,26], and erosion of the lithosphere by a flat-subducting slab[27,28]. However, no tectonic processes of these types are thought to have operated during the mid to late Miocene (<~15 Myr) to present near the Colorado plateau.

As an alternative, we suggest that foundering is driven by thermo-chemical convection (Fig. 4): adiabatically upwelling mantle generates buoyant basaltic partial melts that intrude the base and sides of the Colorado plateau lithosphere. Crystallization of such melts can destabilize the base of the lithosphere by refertilizing the pre-existing depleted mantle with basalt/eclogite/pyroxenite dykes or diffuse metasomatic zones[29]. Peridotite xenoliths in Eocene minette dykes in the central Colorado plateau show such rejuvenation. These xenoliths define a geotherm that indicates a thermochemical boundary layer
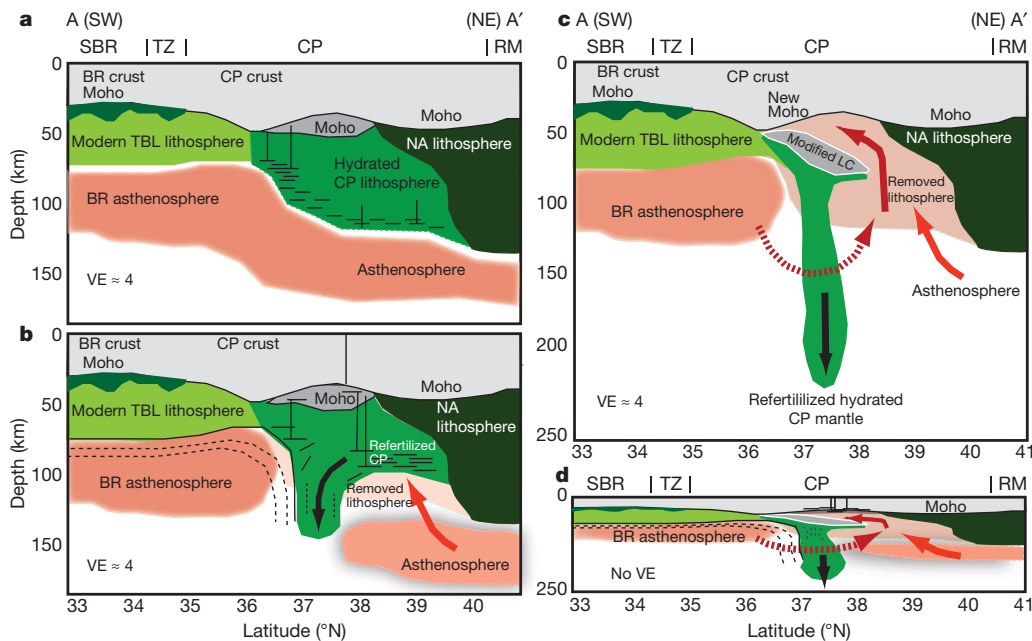
Figure 4 | The inferred progression of events associated with lithospheric foundering and the geometry of the drip today. a, The lithosphere and asthenosphere as they existed at some time following Farallon slab removal. The Colorado plateau lithosphere has been hydrated, and is being invaded by small-volume melts from the asthenosphere. The ~50 km of topography on the LAB along the southwestern edge of the Colorado plateau permits small-volume decompression melting. BR, Basin and Range; NA, North American lithosphere; TBL, thermal boundary layer. **b**, The small increase in density from freezing melts, schematically shown as solid back lines, and the viscosity reduction from hydration and advected heat, destabilizes the lithosphere and initiates a localized downwelling (black arrow and black dashed flow lines). **c, d**, The drip as we infer it from the seismic data, shown at respectively 4:1 and 1:1 vertical exaggeration (VE). The re-fertilized Colorado plateau mantle has been removed almost completely, delaminating the lowermost crust with it. Asthenosphere is invading the region from beneath the drip (solid red arrows) and around the peripheries of the drip (dashed red arrow). We infer that the lowermost crust (LC) involved in the downwelling has been modified by intrusion of basaltic melts that froze to produce high density eclogites. In **a–d**, the range of latitude shown corresponds to ~1,000 km.

thickness of ~120–150 km. The xenoliths show a compositional spectrum (in terms of Mg number, Mg# = atomic Mg/(Mg+Fe)) of 0.91–0.92—typical of depleted cratonic mantle—to values lower than the canonical value of ambient convecting mantle (Mg# = 0.89)[30], suggesting that the base of the lithosphere may have been substantially refertilized. Refertilization leads to a small but important compositional density increase that can exceed the original positive buoyancy of depleted Colorado plateau cratonic mantle. Refertilization by melt infiltration also would have advected heat into the base of a North American lithosphere. Already hydrated by low-angle Farallon plate subduction, it would be further weakened thermally, reducing viscous resistance to the newly acquired negative buoyancy forces and thereby initiating foundering.

We suggest that such magmatism commenced with mantle return flow in response to late Eocene–early Miocene rollback or collapse of the flat-subducting Farallon plate. This period was characterized by an ignimbrite flare-up[31], manifested in the formation of numerous laccoliths, plutons and porphyry copper deposits in western North America[32]. Any magma-induced destabilization mechanism is ultimately driven by the magmatic flux, which depends on the amount of 'head-space' for adiabatic upwelling of the underlying undepleted mantle, making the pre-existing and evolving topography of the LAB a primary control on such processes (Fig. 4). For example, thin lithosphere adjacent to thick lithosphere would facilitate generation of more magmas and hence be more prone to thermochemical destabilization. The thick lithosphere of the Colorado plateau was resistant to such destabilization, but was vulnerable to destabilization around its margins from the sides where it abuts thin, extended Basin and Range lithosphere. This thermochemical erosion will always initiate wherever the LAB rapidly shallows. As the drip progresses, more head-space is created for melt invasion and refertilization on its inboard side, further feeding the instability (Fig. 4). Ultimately, the entire thickness of lithospheric mantle is involved in the instability, which reaches the lower

crust: overall, the deeper foundering appears as a convective drip, but at the lower crust, which is relatively weak in comparison to the uppermost mantle, the downwelling takes the form of a delamination-like process, with deformation guided by the strength contrast between upper mantle and lower crust.

We propose that a series of such events have been removing the lithosphere from the Colorado plateau peripheries since the Farallon slab was removed 20–30 Myr ago, and that we have imaged only the most recent of these. These events are responsible for the uneven, outside-in magmatic invasion of the plateau, as well as uplift of its edges and interior.

METHODS SUMMARY

Finite-frequency seismic body wave tomography is a standard means of including Fresnel zone effects in constructing a three-dimensional image of the Earth from teleseismic travel times. The finite-frequency Rayleigh wave tomography is a regional study of the Colorado plateau and its surroundings. This form of Rayleigh wave tomography gives absolute isotropic v_S measurements, and, because the waves travel horizontally, provides somewhat better vertical resolution of upper mantle v_S structure than teleseismic body wave tomography. Rayleigh waves in the band we examined (0.0067–0.05 Hz) are sensitive to v_S structure from the mid-crust, ~20 km depth, into the upper mantle, ~150–200 km depth.

Receiver function imaging is a means to make a scattered wave image of the subsurface using P to S (or S to P) converted waves, usually from teleseismic earthquake signals. The PdS receiver function is created by deconvolving the direct P wave recorded on the vertical component of motion from the radial component of motion. The deconvolution removes the earthquake source function, and replaces it with a known shaping pulse. The time difference between direct P and the scattered S waves on the receiver function, combined with the incidence angle of the P wave and an estimated velocity model, are used to back-project the receiver function in space to the points of conversion from P to S. Summing partial images for many earthquakes recorded by many stations improves the signal to noise ratio of the image. The process for making SdP receiver functions is similar. Direct teleseismic S waves have lower frequency (<0.2–0.1 Hz) content than the teleseismic P waves (<~2.0 Hz), and therefore produce a lower frequency image.

1. Thompson, G. A. & Zoback, M. L. Regional geophysics of the Colorado Plateau. *Tectonophysics* **61,** 149–181 (1979).
2. Schmandt, B. & Humphreys, E. D. Complex subduction and small-scale convection revealed by body-wave tomography of the western United States upper mantle. *Earth Planet. Sci. Lett.* **297,** 435–445 (2010).
3. Humphreys, E. *et al.* How Laramide-age hydration of North American lithosphere by the Farallon slab controlled subsequent activity in the western United States. *Int. Geol. Rev.* **45,** 575–595 (2003).
4. Lee, C.-T. A., Yin, Q., Rudnick, R. L. & Jacobsen, S. B. Preservation of ancient and fertile lithospheric mantle beneath the southwestern United States. *Nature* **411,** 69–73 (2001).
5. Li, Z.-X. A., Lee, C.-T., Peslier, A. H., Lenardic, A. & Mackwell, S. J. Water contents in mantle xenoliths from the Colorado Plateau and vicinity: implications for the mantle rheology and hydration-induced thinning of continental lithosphere. *J. Geophys. Res.* **113,** B09210, doi:10.1029/2007JB005540 (2008).
6. Roy, M., Jordan, T. H. & Pederson, J. Colorado Plateau magmatism and uplift by warming of heterogeneous lithosphere. *Nature* **459,** 978–982 (2009).
7. Crow, R. S. *et al.* Shrinking of the Colorado Plateau via lithospheric mantle erosion: evidence from Nd and Sr isotopes and geochronology of Neogene basalts. *Geology* **39,** 27–30 (2011).
8. Karlstrom, K. E., Crow, R., Crossey, L. J., Coblentz, D. & Van Wijk, J. W. Model for tectonically driven incision of the younger than 6 Ma Grand Canyon. *Geology* **36,** 835–838 (2008).
9. Smith, D. Insights into the evolution of the uppermost continental mantle from xenolith localities on and near the Colorado Plateau and regional comparisons. *J. Geophys. Res.* **105,** 16769–16781 (2000).
10. Livaccari, R. F. & Perry, F. V. Isotopic evidence for preservation of Cordilleran lithospheric mantle during the Sevier-Laramie orogeny, western United States. *Geology* **21,** 719–722 (1993).
11. Humphreys, E. D. Post-Laramide removal of the Farallon slab, western United States. *Geology* **23,** 987–990 (1995).
12. Huntington, K. W., Wernicke, B. P. & Eiler, J. M. Influence of climate change and uplift on Colorado Plateau paleotemperatures from carbonate clumped isotope thermometry. *Tectonics* **29,** TC3005, doi:10.1029/2009TC002449 (2010).
13. Spencer, J. Uplift of the Colorado Plateau due to lithosphere attenuation during Laramide low-angle subduction. *J. Geophys. Res.* **101,** 13595–13609 (1996).
14. Moucha, R. *et al.* Deep mantle forces and the uplift of the Colorado Plateau. *Geophys. Res. Lett.* **36,** L19310, doi:10.1029/2009GL039778 (2009).
15. van Wijk, J. W. *et al.* Small-scale convection at the edge of the Colorado Plateau: implications for topography, magmatism, and evolution of Proterozoic lithosphere. *Geology* **38,** 611–614 (2010).
16. Benz, H. M. & McCarthy, J. Evidence for an upper mantle low velocity zone beneath the southern Basin and Range – Colorado Plateau transition zone. *Geophys. Res. Lett.* **21,** 509–512 (1994).
17. Henstock, T. J., Levander, A. & DeepProbe Working Group. Probing the Archean and Proterozoic lithosphere of western North America. *GSA Today* 8, 1–5 16–17 (1998).
18. Buehler, J. & Shearer, P. Pn tomography of the western United States using USArray. *J. Geophys. Res.* 115 B09315 (2010).
19. Schmandt, B. & Humphreys, E. Complex subduction and small-scale convection revealed by body-wave tomography of the western United States upper mantle. *Earth Planet. Sci. Lett.* **297,** 435–445 (2010).
20. Bassin, C., Laske, G. & Masters, G. The current limits of resolution for surface wave tomography in North America. *Eos* **81,** F897 (2000).
21. Bird, P. Continental delamination and the Colorado Plateau. *J. Geophys. Res.* **84,** 7561–7571 (1979).
22. Morency, C. & Doin, M.-P. Numerical simulations of the mantle lithosphere delamination. *J. Geophys. Res.* **109,** B03410, doi:10.1029/2003JB002414 (2004).
23. Conrad, C. P. & Molnar, P. Convective instability of a boundary layer with temperature- and strain-dependent viscosity in terms of 'available buoyancy'. *Geophys. J. Int.* **139,** 51–68 (1999).
24. Bird, P. Laramide crustal thickening event in the Rocky Mountain foreland and Great Plains. *Tectonics* **3,** 741–758 (1984).
25. Kay, S. M., Coira, B. & Viramonte, J. Young mafic back arc volcanic rocks as indicators of continental lithospheric delamination beneath the Argentine Puna Plateau, Central Andes. *J. Geophys. Res.* **99,** 24323–24339 (1994).
26. Ducea, M. N. & Saleeby, J. B. Buoyancy sources for a large, unrooted mountain range, the Sierra Nevada, California; evidence from xenolith thermobarometry. *J. Geophys. Res.* **101,** 8229–8244 (1996).
27. Bird, P. Formation of the Rocky Mountains, Western United States: a continuum computer model. *Science* **239,** 1507 (1988).
28. Menzies, M., Xu, Y., Zhang, H. & Fan, W. Integration of geology, geophysics and geochemistry: a key to understanding the North China Craton. *Lithos* **96,** 1–21 (2007).
29. Foley, S. F. *et al.* The composition of near-solidus melts of peridotite in the presence of CO_2 and H_2O between 40 and 60 kbar. *Lithos* **112,** 274–283 (2009).
30. Ehrenberg, S. N. Petrogenesis of garnet lherzolite and megacrystalline nodules from the Thumb, Navajo volcanic field. *J. Petrol.* **23,** 507–547 (1982).
31. Lipman, P. W. in *The Cordilleran Orogen: Conterminous U. S.* (eds Burchfiel, B. C., Lipman, P. W. & Zoback, M. L.) 481–514 (Geological Society of America, 1992).
32. Nelson, S. T., Davidson, J. P. & Sullivan, K. R. New age determinations of central Colorado Plateau Laccoliths, Utah — recognizing disturbed K-Ar systematics and re-evaluating tectonomagmatic relationships. *Geol. Soc. Am. Bull.* **104,** 1547–1560 (1992).

METHODS

The finite-frequency body wave tomography is described in a study of southern California[33]. Finite-frequency seismic body wave tomography is a standard means of including Fresnel zone effects in constructing a three-dimensional image of the Earth from teleseismic travel times[2]. The finite-frequency Rayleigh wave tomography is a regional study of the Colorado plateau[34] and its surroundings, calculated with the two-plane wave method[35]. This form of Rayleigh wave tomography gives absolute isotropic v_S measurements, and, because the waves travel horizontally, provides somewhat better vertical resolution of upper-mantle v_S structure than teleseismic body wave tomography. Rayleigh waves in the band we examined (0.0067–0.05 Hz) are sensitive to v_S structure from the mid-crust ($\sim$20 km) into the upper mantle ($\sim$150–200 km depth). Rayleigh waves are insensitive to v_P and density structure at these depths.

Receiver function imaging is a means to make a scattered wave image of the subsurface using P to S (or S to P) converted waves[36,37], usually from earthquake signals. The Ps receiver function is created by deconvolving the direct P wave on the vertical (or rotated longitudinal) component of motion from the radial (or rotated S) component and the tangential component. The deconvolution removes the earthquake source function and replaces it with a known shaping pulse. The time difference between direct P and the scattered S waves on the receiver function (or direct S and scattered P), combined with a known or estimated incidence angle and back-azimuth of the P-wave, can be used to back-project the receiver function laterally and in depth to the point of conversion from P to S. The back projection requires an estimate of the P and S velocity fields. Summing many partial images from receiver functions recorded at many seismograph stations for many earthquakes, a process called common conversion point (CCP) stacking[38], improves the signal to noise ratio of the image. For Ps receiver functions, earthquakes between $\sim$40° and 90° from the seismograph station are used. The process for making SdP receiver functions is similar[37]: The direct S arrival is deconvolved from the longitudinal component to isolate scattered P waves. Earthquakes between 55° and 85° from the seismograph station are used. The direct teleseismic S waves have lower frequency ($<$0.2–0.1 Hz) content than the teleseismic P waves ($<\sim$2.0 Hz) owing to higher attenuation in the Earth, and therefore produce a lower frequency image.

33. Schmandt, B. & Humphreys, E. Seismic heterogeneity and small-scale convection in the southern California upper mantle. *Geochem. Geophys. Geosyst.* **11,** Q05004, doi:10.1029/2010GC003042 (2010).
34. Liu, K., Levander, A., Niu, F. & Miller, M. S. Imaging crustal and upper mantle structure beneath the Colorado Plateau using finite-frequency Rayleigh wave tomography. *Geochem. Geophys. Geosyst.* (submitted).
35. Forsyth, D. W. & Li, A. in *Seismic Earth: Array Analysis of Broadband Seismograms* Vol. 157 (eds Levander, A. & Nolet, G.) 81–97 (American Geophysical Union, 2005).
36. Langston, C. A. Corvallis, Oregon, crustal and upper mantle receiver structure from teleseismic P and S waves. *Bull. Seismol. Soc. Am.* **67,** 713–724 (1977).
37. Yuan, X., Kind, R., Li, X. & Wang, R. The S receiver functions: synthetics and data example. *Geophys. J. Int.* **165,** 555–564 (2006).
38. Dueker, K. G. & Sheehan, A. F. Mantle discontinuity structure form midpoint stacks of converted P to S waves across the Yellowstone hotspot track. *J. Geophys. Res.* **102,** 8313–8327 (1997).

Activation of the innate immune receptor Dectin-1 upon formation of a 'phagocytic synapse'

Helen S. Goodridge[1,2,3], Christopher N. Reyes[1], Courtney A. Becker[1], Tamiko R. Katsumoto[4], Jun Ma[1], Andrea J. Wolf[1], Nandita Bose[5], Anissa S. H. Chan[5], Andrew S. Magee[5], Michael E. Danielson[5], Arthur Weiss[4,6], John P. Vasilakos[5] & David M. Underhill[1,3]

Innate immune cells must be able to distinguish between direct binding to microbes and detection of components shed from the surface of microbes located at a distance. Dectin-1 (also known as CLEC7A) is a pattern-recognition receptor expressed by myeloid phagocytes (macrophages, dendritic cells and neutrophils) that detects β-glucans in fungal cell walls and triggers direct cellular antimicrobial activity, including phagocytosis and production of reactive oxygen species (ROS)[1,2]. In contrast to inflammatory responses stimulated upon detection of soluble ligands by other pattern-recognition receptors, such as Toll-like receptors (TLRs), these responses are only useful when a cell comes into direct contact with a microbe and must not be spuriously activated by soluble stimuli. In this study we show that, despite its ability to bind both soluble and particulate β-glucan polymers, Dectin-1 signalling is only activated by particulate β-glucans, which cluster the receptor in synapse-like structures from which regulatory tyrosine phosphatases CD45 and CD148 (also known as PTPRC and PTPRJ, respectively) are excluded (Supplementary Fig. 1). The 'phagocytic synapse' now provides a model mechanism by which innate immune receptors can distinguish direct microbial contact from detection of microbes at a distance, thereby initiating direct cellular antimicrobial responses only when they are required.

Studies in mice and humans have demonstrated an important role for Dectin-1 in anti-fungal defence[3–6]. Dectin-1 signals activate antimicrobial (phagocytosis, production of ROS) and inflammatory (cytokine and chemokine production) innate immune responses, and influence the development of adaptive immunity (reviewed in refs 1, 2). Although Dectin-1 has been demonstrated to collaborate with TLR signals to orchestrate immune responses to fungi[7,8], it activates a distinctly different signalling cascade. Dectin-1 signals via a motif in its cytoplasmic tail that resembles an immunoreceptor tyrosine-based activation motif (ITAM; reviewed in refs 1, 2). Like other ITAM-based receptors, including Fc receptors (FcR), T cell receptors (TCR), and B cell receptors (BCR), Dectin-1 signalling relies on activation of Src and Syk family kinases. However, in contrast to conventional ITAMs which comprise dual YXXL sequences, Dectin-1's "hemITAM" has only a single YXXL[2,9]. Despite its unusual ITAM, Dectin-1 ligation by β-glucan-containing particles, such as zymosan and curdlan, triggers Src/Syk-dependent downstream signals in myeloid cells (macrophages, dendritic cells (DC) and neutrophils) to activate mitogen-activated protein (MAP) kinases, as well as NF-κB and NFAT transcription factors (reviewed in refs 1, 2). In addition to inflammatory responses that are also triggered by TLRs, Dectin-1 induces distinct antimicrobial responses. Dectin-1 is a key phagocytic receptor for fungi and triggers a massive oxidative burst in response to fungal exposure[5,6,10,11].

TLRs sense soluble microbial stimuli and are activated by dimerization of intracellular signalling domains. The decades-old use of the small (6–8 kDa), soluble β-glucan laminarin (from *Laminaria digitata*) to 'block' β-glucan receptors on macrophages rather than activate them suggests that Dectin-1 may behave very differently[12]. Indeed, in previous studies we have failed to detect Src, Syk and NFAT activation following treatment of macrophages with laminarin even though the material is a polymer of pure ligand[11,13]. We proposed that a larger molecule may be required to provide a greater degree of receptor cross-linking to permit activation. We therefore compared the ability of whole glucan particles (WGP), a particulate *Saccharomyces cerevisiae* β-glucan preparation that lacks TLR-stimulating activity (Supplementary Fig. 2), with various molecular weight fractions of soluble *S. cerevisiae* β-1,3/1,6-glucans (see Fig. 1a).

WGP, like zymosan and curdlan, induced robust Dectin-1-dependent responses, including phagocytosis, induction of TNF-α, IL-6 and ROS by bone marrow-derived macrophages and DC (BMM and BMDC; Fig. 1b, c and Supplementary Figs 3–5, and data not shown). In contrast, none of the soluble β-glucans, not even the high molecular weight fraction, induced ROS, TNF-α or IL-6 production by either BMM or BMDC. Similar data were obtained using murine neutrophils and human monocytes and monocyte-derived macrophages (Supplementary Figs 4 and 6, and data not shown). Like zymosan, WGP induced Dectin-1 signalling (activation of Syk, p38 MAP kinase, NF-κB and NFAT; Fig. 1d, e and Supplementary Figs 5d and 7–9). In contrast, none of the soluble β-glucans induced Dectin-1 signals. Thus simply increasing the size of the β-glucan polymer is not sufficient to activate Dectin-1 signalling.

It was demonstrated previously that a 150 kDa soluble *S. cerevisiae* β-glucan interacts with purified Dectin-1 with picomolar affinity[14]. We used a variety of approaches to verify that our *S. cerevisiae* β-glucans bind directly to cell surface Dectin-1. Fluorescently labelled soluble β-glucans bound to BMM and BMDC surfaces in a Dectin-1-dependent manner and all molecular weight fractions efficiently blocked (by at least 50%) binding of anti-Dectin-1 antibodies at the dose ($50 \, \mu g \, ml^{-1}$) used in this study (Fig. 2a and Supplementary Figs 10 and 11, and data not shown). Furthermore, we observed significant soluble β-glucan binding to Dectin-1-expressing RAW264.7 macrophages, which, like primary macrophages/DC, respond robustly to particulate but not soluble β-glucans (Supplementary Figs 12 and 13). In contrast, parental RAW264.7 macrophages, which express very little Dectin-1 and mount only low responses to β-glucan particles, failed to bind soluble β-glucans (Supplementary Fig. 12a and data not shown). Furthermore, like laminarin, all the soluble β-glucans blocked Dectin-1-mediated particulate β-glucan responses in primary BMM and BMDC (Supplementary Figs 14 and 15). Thus, despite efficient binding to Dectin-1, soluble β-glucans are incapable of activating the receptor.

[1]IBD and Immunobiology Research Institute, 8700 Beverly Boulevard, Cedars-Sinai Medical Center, Los Angeles, California 90048, USA. [2]Regenerative Medicine Institute, Cedars-Sinai Medical Center, 8700 Beverly Boulevard, Los Angeles, California 90048, USA. [3]David Geffen School of Medicine at UCLA, 10833 Le Conte Avenue, University of California, Los Angeles, California 90095, USA. [4]Department of Medicine, Rosalind Russell Medical Research Center for Arthritis, 513 Parnassus, University of California, San Francisco, California 94143, USA. [5]Biothera, 3388 Mike Collins Drive, Eagan, Minnesota 55121, USA. [6]Howard Hughes Medical Institute, University of California, San Francisco, California 94143, USA.
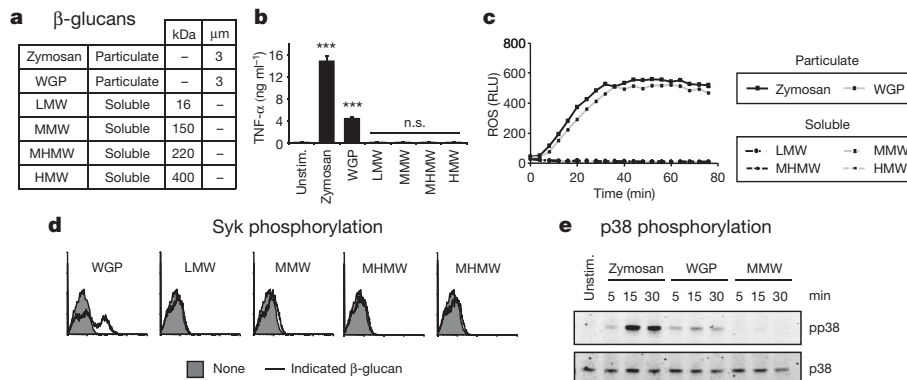
Figure 1 | Particulate, but not soluble, β-glucans induce Dectin-1 signalling. **a**, Size (molecular weight or diameter) of β-glucan preparations used in this study. **b–e**, Bone marrow-derived macrophages (BMM; **b–d**, IFN-γ-primed overnight) were stimulated with 50 μg ml⁻¹ β-glucans. **b**, TNF-α production (24 h) was assessed by ELISA; data are means plus s.d. of triplicate culture (***$P < 0.001$; n.s., not significant). Unstim., unstimulated. **c**, ROS production was assessed by luminol-enhanced chemiluminescence (ECL); data points are means of triplicate culture. RLU, relative light units. **d**, Syk activation (10 min) was assessed by intracellular flow cytometry. **e**, p38 MAP kinase activation at the indicated times was assessed by immunoblotting. All data are representative of at least three independent experiments.

These data indicate that the mode of presentation of the β-glucan may be critical for Dectin-1 signalling. We therefore examined whether immobilization of soluble β-glucans is sufficient to trigger Dectin-1 signalling. Like the β-glucan particles, soluble β-glucans immobilized on tissue culture plates or polystyrene latex beads (0.5 μm diameter or larger) stimulated robust Dectin-1-dependent responses (Fig. 2b–e and Supplementary Figs 16–18).

The above data collectively demonstrate that to activate Dectin-1, β-glucans must be presented in an immobilized form, for example, on the surface of a phagocytosable particle such as a yeast cell. This scenario is reminiscent of the requirement for antigen presentation to the TCR by an antigen-presenting cell (APC). TCR signalling is regulated by CD45, a membrane protein with a large extracellular domain and intrinsic tyrosine phosphatase activity[15,16]. CD45 is initially required for removal of an inhibitory phosphate to permit activation of Src family kinases, but subsequently must be isolated from the TCR complex due to its negative regulation of ITAM signalling. We therefore investigated whether Dectin-1 signalling is similarly regulated by CD45

and/or CD148, a CD45-related membrane tyrosine phosphatase with overlapping function that regulates ITAM signalling by the TCR, BCR and FcγR[17,18].

CD45-deficient BMM exhibited normal zymosan phagocytosis and only partially compromised WGP-induced ROS and TNF-α production, whereas CD148-deficient BMM showed no defect (Fig. 3a–c). In contrast, these Dectin-1 responses were severely compromised in BMM deficient in both CD45 and CD148 (Fig. 3a–c). TNF-α induction by various other stimuli was unaffected (Supplementary Fig. 19a) and Dectin-1 surface expression was normal (Supplementary Fig. 19b). WGP failed to induce Syk activation in BMM deficient in both phosphatases (Fig. 3d), which as previously reported[18] had elevated basal levels of phosphorylation of the Src family kinase Lyn at its inhibitory tyrosine (Y507; Fig. 3e). Thus CD45 and CD148 have overlapping function in the regulation of Dectin-1 signalling in macrophages.

TCR activation is characterized by formation of an 'immunological synapse' between a T cell and an APC. Surface molecules on the two interacting cells are reorganized at the cell–cell interface to permit TCR
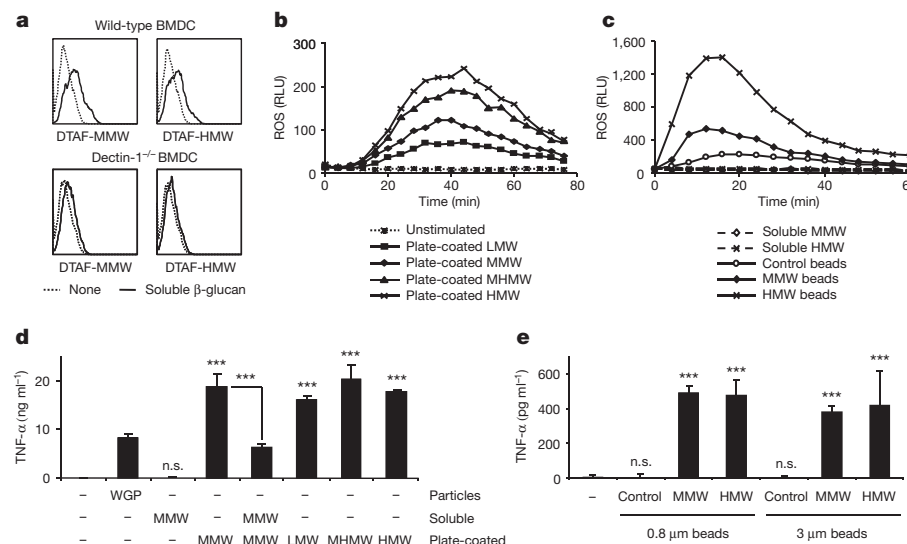

Figure 2 | Immobilized β-glucans induce Dectin-1 signalling. **a**, Soluble β-glucan binding (50 μg ml⁻¹, 10 min) to wild type and Dectin-1⁻/⁻ BMDC was assessed by flow cytometry. **b**, **c**, IFN-γ-primed BMM were stimulated with soluble β-glucans (50 μg ml⁻¹) or β-glucans immobilized on either tissue culture plates (**b**) or 0.8 μm polystyrene latex beads (**c**). ROS production was measured by luminol-ECL; data points are means of triplicate culture. **d**, **e**, TNF-α production (24 h) by BMDC exposed to soluble/particulate (50 μg ml⁻¹), plate-immobilized or bead-coated soluble β-glucans was assessed by ELISA; data are expressed as means plus s.d. of triplicate culture (***$P < 0.001$; n.s., not significant). All data are representative of at least three independent experiments.
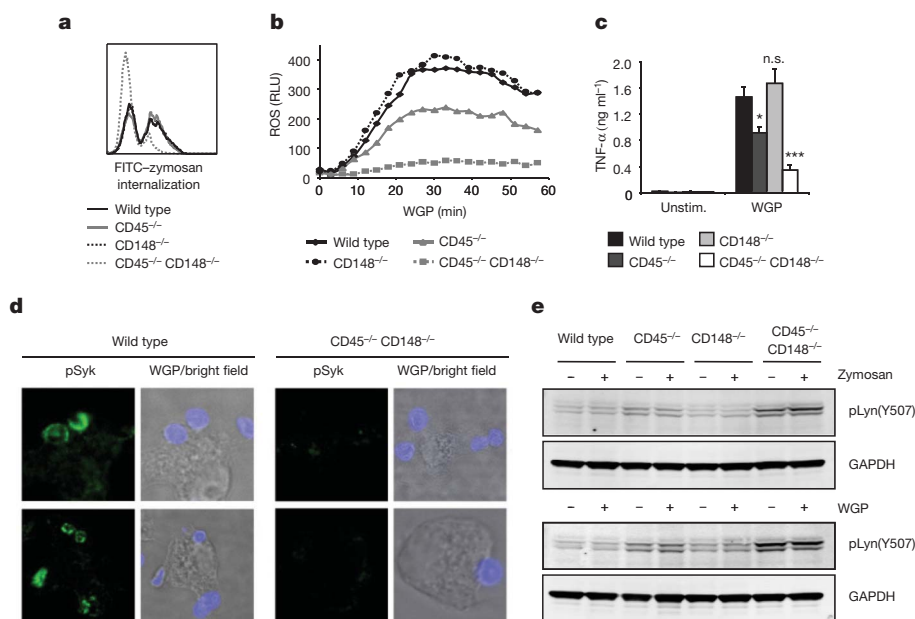
Figure 3 | CD45 and CD148 regulate Dectin-1 signalling. Responses of wild-type and CD45/CD148-deficient BMM (**b**, IFN-γ-primed) were examined. **a**, Internalization of FITC-labelled zymosan particles (20 μg ml⁻¹, 10 min) was assessed by flow cytometry. **b**, ROS production was measured by luminol-ECL; data points are means of triplicate culture. **c**, TNF-α production (50 μg ml⁻¹ WGP, 24 h) was assessed by ELISA; data are expressed as means plus s.d. of triplicate culture (*P < 0.01; ***P < 0.001; n.s., not significant). **d**, Syk activation (pSyk; green) by AlexaFluor647-labelled WGP (20 μg ml⁻¹, 1 min; blue) was assessed by confocal microscopy. **e**, Inactive Lyn (pY507) levels following zymosan/WGP stimulation (50 μg ml⁻¹, 10 min) were assessed by immunoblotting. All data are representative of three independent experiments.

signalling[15]. Following the initial CD45-promoted activation of Src family kinases, CD45 must be isolated from the TCR complex in order to remove its inhibitory phosphatase activity and permit propagation of the ITAM signal. We proposed that following binding of β-glucan-containing particles, CD45 and CD148 would similarly be sequestered from Dectin-1 to permit activation of Dectin-1 signalling.

We therefore examined macrophage surface molecule rearrangement following β-glucan particle binding. As we and others have observed previously[8,19], Dectin-1 clustered at β-glucan particle contact sites on the surface of Dectin-1-expressing RAW264.7 macrophages and phagocytic cups formed within 1 min of binding (Fig. 4a and Supplementary Fig. 20a). Recruitment of active Src and Syk kinases,
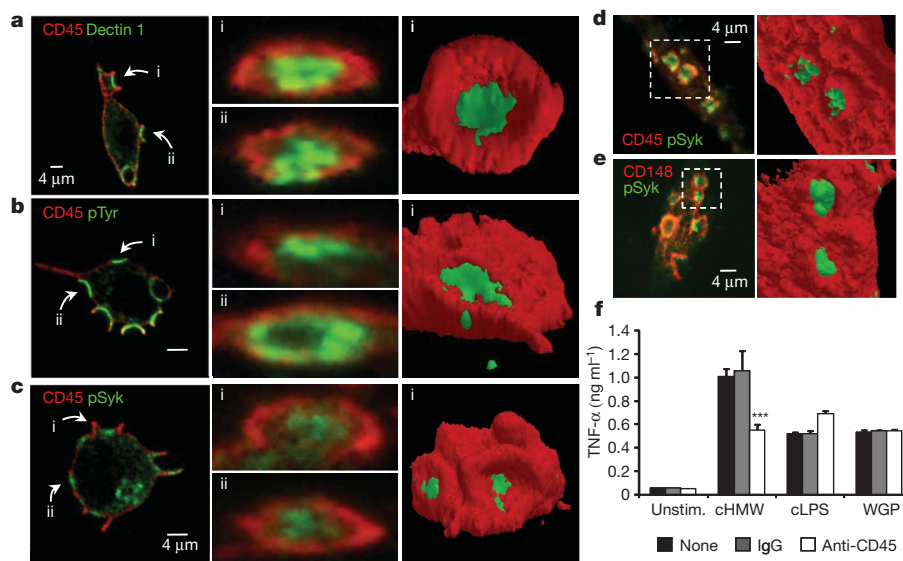

Figure 4 | CD45 and CD148 phosphatases are excluded from the β-glucan particle contact site. a–c, Confocal microscopy of SBPc-tagged Dectin-1-expressing RAW264.7 macrophages stimulated with zymosan (20 μg ml⁻¹, 1 min) and stained for CD45 (red) and SBPc tag (Dectin-1; green), phospho-tyrosine (pTyr; green) or active Syk (pSyk; green). z-stacks were analysed to visualize the indicated particle contact sites (left panels) in cross-section (centre panels). Three-dimensional isosurface models of the indicated contact sites (right panels) were generated using ImageJ and ImageSurfer. Scale bar = 4 μm. **d**, Resident peritoneal macrophages stimulated with WGP (20 μg ml⁻¹, 1 min) were stained (left panel) for CD45 (red) and active Syk (pSyk; green). **e**, Dectin-1-expressing RAW264.7 macrophages stimulated with WGP (20 μg ml⁻¹, 1 min) were stained (left panel) for CD148 (red) and active Syk (pSyk; green). Isosurface models (**d, e**, right panels) are of the indicated particle contact sites. **f**, BMDC were added to tissue culture plates pre-coated with HMW soluble β-glucan (cHMW) or LPS (cLPS) and/or anti-CD45 or control IgG; some BMDC were stimulated with WGP (50 μg ml⁻¹). TNF-α production (24 h) was assessed by ELISA; data are means plus s.d. of triplicate culture (***P < 0.001). All data are representative of at least three independent experiments.

as well as other tyrosine-phosphorylated proteins, to the contact sites of β-glucan particles with macrophages (Dectin-1-expressing RAW264.7 and primary murine macrophages) was also seen at this early time point (Fig. 4b–e and Supplementary Figs 20b–c and 21–23).

In contrast, CD45 was uniformly distributed on the surface of unstimulated macrophages, but upon contact with β-glucan particles was specifically absent from regions of membrane enriched for Dectin-1, active Src family and Syk kinases, and phospho-tyrosine staining (Fig. 4a–d and Supplementary Figs 20–24). Time course experiments showed that CD45 exclusion from the region of Dectin-1 clustering occurs before initiation of phagocytic cup formation (Supplementary Fig. 25 and Supplementary Movie 1), and is not dependent on actin dynamics (Supplementary Fig. 26). CD148 was similarly excluded from regions of Dectin-1 clustering and signalling induced by β-glucan particle binding (Fig. 4e). Three-dimensional images clearly show a 'bulls-eye' pattern of staining with a central Dectin-1 signalling region, from which CD45 and CD148 are excluded (Fig. 4a–e, Supplementary Fig. 27 and Supplementary Movie 2). We also observed clustering of Dectin-1 and exclusion of CD45 upon contact of Dectin-1-expressing RAW264.7 macrophages with live fungi (*S. cerevisiae*, *Candida albicans* and *Aspergillus fumigatus*) and β-glucan-coated plates (Supplementary Figs 28–32). In contrast, upon binding of soluble β-glucans to Dectin-1, CD45 remained colocalized with Dectin-1 at the cell surface (Supplementary Fig. 33).

CD45 and CD148 exclusion from the contact site of Dectin-1 with the β-glucan particle in the forming phagosome (or the contact surface during frustrated phagocytosis) is consistent with the hypothesis that following the initial activation of Src family kinases, the phosphatases must be isolated from the receptor to permit productive signalling. This model predicts that co-coating anti-CD45 antibodies on plates with β-glucans to prolong CD45 colocalization with β-glucan-bound Dectin-1 would suppress Dectin-1 signalling. Indeed, co-coated anti-CD45 reduced immobilized β-glucan-induced TNF-α production by about 45%, but did not affect TNF-α induction by WGP (which contact cells at points not bound to anti-CD45) or immobilized lipopolysaccharide (LPS; Fig. 4f).

Collectively, our data suggest a mechanism by which Dectin-1, unlike other innate pattern-recognition receptors such as TLRs, discriminates between soluble and particulate ligands (Supplementary Fig. 1). Binding of particulate β-glucans (such as yeast) to Dectin-1 triggers phagocytosis, a process that involves massive reorganization of membrane proteins and membrane movement coordinated by the actin cytoskeleton. We have shown that during this process membrane tyrosine phosphatases, which are well characterized regulators of ITAM signalling[16], are excluded from the particle contact site. In a manner analogous to the formation of immunological synapses between APCs and T cells, 'phagocytic synapses' are required for activation of Dectin-1 when myeloid cells encounter β-glucan-containing microbes. In contrast, it seems that upon detection of soluble β-glucans, the inhibitory activity of membrane tyrosine phosphatases cannot be sufficiently isolated from the crosslinked receptors, and thus Dectin-1 signalling is aborted.

Future studies are required to determine whether the unique nature of the Dectin-1 hemITAM underlies its requirement for ligand immobilization. We suspect that the C-type lectin family member CLEC1B (also known as CLEC2) may be similarly regulated. CLEC2 also contains a hemITAM and signals in an apparently identical manner to Dectin-1 (ref. 20). In addition to having a role in platelet aggregation, CLEC2 is expressed on the surface of neutrophils and is capable of inducing phagocytosis of anti-CLEC2-coated beads[21]. Interestingly, Dectin-1–CLEC2 chimaeras can be activated to induce TNF-α production by zymosan, but this activation is blocked by soluble β-glucans[21] (which presumably bind but fail to activate the receptor), indicating that activation of CLEC2 signalling may be dependent on the formation of a phagocytic synapse.

It is widely accepted that the nature of an innate immune response to a microbe is defined by the types of pattern-recognition receptors that detect it. Thus receptors that detect viral nucleotides induce responses appropriate for killing viruses. Activation of phagocytic receptors is only appropriate when they bind intact microbes. Although models exist for the detection of soluble stimuli (for example, receptor dimerization, induction of conformational changes), we are currently lacking good models for the discrimination of soluble versus particulate ligands. Our data present the phagocytic synapse as a mechanistic model for the specific detection of ligands associated with a microbial surface, as opposed to those released from distantly located organisms.

METHODS SUMMARY

β-glucan preparations. Particulate *S. cerevisiae* β-glucans (zymosan; Sigma) and whole glucan particles (Wellmune WGP; Biothera) were used as described previously[8,22]. Soluble *S. cerevisiae* β-glucans were prepared by acid hydrolysis of WGP and fractionated by preparative gel permeation chromatography. The molecular weight distribution of each soluble β-glucan was determined by gel permeation chromatography (GPC) with multi-angle laser light scattering photometry (MALLS); the polydispersities of the LMW, MMW, MHMW and HMW (low, medium, medium-high, and high molecular weight, respectively) soluble β-glucans are 1.5, 1.6, 1.2 and 1.4, respectively. All β-glucan preparations were endotoxin-free and used at 50 μg ml^{-1} unless otherwise stated.

Confocal microscopy. Cells were plated on glass coverslips overnight before addition of β-glucan particles, centrifuged briefly to ensure particle contact with the cells, and incubated at 37 °C for the times indicated. Cells were washed to remove unbound particles, fixed with 10% formalin, permeabilized with ice-cold acetone, blocked and stained with primary and secondary antibodies. Coverslips were mounted and examined using a Leica TCS SP5 confocal microscope. Image analysis was performed using Leica LAS AF software, as well as ImageJ and ImageSurfer[23].

β-glucan immobilization and CD45 co-immobilization. Soluble β-glucans were immobilized on tissue culture plates or polystyrene latex beads by incubation with PBS/EDTA containing 100 μg ml^{-1} soluble β-glucan for 1 h at 37 °C. Plates or beads were then washed to remove unbound β-glucans, and blocked with media containing 10% fetal calf serum before use. For CD45 co-immobilization assays, HMW soluble β-glucan (20 μg ml^{-1}) and LPS (100 ng ml^{-1}) were immobilized on tissue culture plates in PBS/EDTA in the presence or absence of 10 ng ml^{-1} anti-CD45 or rat IgG for 1 h at 37 °C. Plates were washed and blocked as above before addition of BMDC.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Goodridge, H. S., Wolf, A. J. & Underhill, D. M. Beta-glucan recognition by the innate immune system. *Immunol. Rev.* **230**, 38–50 (2009).
2. Kerrigan, A. M. & Brown, G. D. Syk-coupled C-type lectin receptors that mediate cellular activation via single tyrosine based activation motifs. *Immunol. Rev.* **234**, 335–352 (2010).
3. Ferwerda, B. *et al.* Human dectin-1 deficiency and mucocutaneous fungal infections. *N. Engl. J. Med.* **361**, 1760–1767 (2009).
4. Plantinga, T. S. *et al.* Early stop polymorphism in human DECTIN-1 is associated with increased *Candida* colonization in hematopoietic stem cell transplant recipients. *Clin. Infect. Dis.* **49**, 724–732 (2009).
5. Saijo, S. *et al.* Dectin-1 is required for host defense against *Pneumocystis carinii* but not against *Candida albicans*. *Nature Immunol.* **8**, 39–46 (2006).
6. Taylor, P. R. *et al.* Dectin-1 is required for β-glucan recognition and control of fungal infection. *Nature Immunol.* **8**, 31–38 (2007).
7. Dennehy, K. M. *et al.* Syk kinase is required for collaborative cytokine production induced through Dectin-1 and Toll-like receptors. *Eur. J. Immunol.* **38**, 500–506 (2008).
8. Gantner, B. N., Simmons, R. M., Canavera, S. J., Akira, S. & Underhill, D. M. Collaborative induction of inflammatory responses by Dectin-1 and Toll-like receptor 2. *J. Exp. Med.* **197**, 1107–1117 (2003).
9. Robinson, M. J., Sancho, D., Slack, E. C., LeibundGut-Landmann, S. & Reis e Sousa, C. Myeloid C-type lectins in innate immunity. *Nature Immunol.* **7**, 1258–1265 (2006).
10. Brown, G. D. *et al.* Dectin-1 is a major β-glucan receptor on macrophages. *J. Exp. Med.* **196**, 407–412 (2002).
11. Underhill, D. M., Rossnagle, E., Lowell, C. A. & Simmons, R. M. Dectin-1 activates Syk tyrosine kinase in a dynamic subset of macrophages for reactive oxygen production. *Blood* **106**, 2543–2550 (2005).
12. Czop, J. K. & Austen, K. F. A beta-glucan inhibitable receptor on human monocytes: its identity with the phagocytic receptor for particulate activators of the alternative complement pathway. *J. Immunol.* **134**, 2588–2593 (1985).

13. Goodridge, H. S., Simmons, R. M. & Underhill, D. M. Dectin-1 stimulation by *Candida albicans* yeast or zymosan triggers NFAT activation in macrophages and dendritic cells. *J. Immunol.* **178,** 3107–3115 (2007).
14. Adams, E. L. *et al.* Differential high-affinity interaction of Dectin-1 with natural or synthetic glucans is dependent upon primary structure and is influenced by polymer chain length and side-chain branching. *J. Pharmacol. Exp. Ther.* **325,** 115–123 (2008).
15. Fooksman, D. R. *et al.* Functional anatomy of T cell activation and synapse formation. *Annu. Rev. Immunol.* **28,** 79–105 (2010).
16. Hermiston, M. L., Zikherman, J. & Zhu, J. W. CD45, CD148, and Lyp/Pep: critical phosphatases regulating Src family kinase signalling networks in immune cells. *Immunol. Rev.* **228,** 288–311 (2009).
17. Lin, J. & Weiss, A. The tyrosine phosphatase CD148 is excluded from the immunologic synapse and down-regulates prolonged T cell signalling. *J. Cell Biol.* **162,** 673–682 (2003).
18. Zhu, J. W., Brdicka, T., Katsumoto, T. R., Lin, J. & Weiss, A. Structurally distinct phosphatases CD45 and CD148 both regulate B cell and macrophage immunoreceptor signalling. *Immunity* **28,** 183–196 (2008).
19. Brown, G. D. *et al.* Dectin-1 mediates the biological effects of β-glucans. *J. Exp. Med.* **197,** 1119–1124 (2003).
20. Fuller, G. L. *et al.* The C-type lectin receptors CLEC-2 and Dectin-1, but not DC-SIGN, signal via a novel YXXL-dependent signalling cascade. *J. Biol. Chem.* **282,** 12397–12409 (2007).
21. Kerrigan, A. M. *et al.* CLEC-2 is a phagocytic activation receptor expressed on murine peripheral blood neutrophils. *J. Immunol.* **182,** 4150–4157 (2009).
22. Goodridge, H. S. *et al.* Differential use of CARD9 by dectin-1 in macrophages and dendritic cells. *J. Immunol.* **182,** 1146–1154 (2009).
23. Feng, D. *et al.* Stepping into the third dimension. *J. Neurosci.* **27,** 12757–12760 (2007).

METHODS

β-glucan preparations. Particulate *S. cerevisiae* β-glucans (zymosan; Sigma) and WGP (Wellmune WGP; Biothera) were used as described previously[8,22]. Soluble *S. cerevisiae* β-glucans were prepared by acid hydrolysis of WGP and fractionated by preparative gel permeation chromatography. The molecular weight distribution of each soluble β-glucan was determined by gel permeation chromatography (GPC) with multi-angle laser light scattering photometry (MALLS); the polydispersities of the LMW, MMW, MHMW and HMW soluble β-glucans were 1.5, 1.6, 1.2 and 1.4, respectively. All β-glucan preparations were endotoxin-free and used at 50 μg ml^{-1} unless otherwise stated.

Live yeast. Live *Saccharomyces cerevisiae* and *Candida albicans* yeast were grown in Sabouraud Dextrose Broth. *Aspergillus fumigatus* conidia were prepared from mature colonies grown on potato dextrose agar by flushing with PBS containing 0.05% Tween-80, and incubated in RPMI for 4 h to generate swollen conidia or 12 h to induce germ tube formation.

Cell culture and functional and biochemical assays. RAW264.7 macrophages stably expressing SBPc-tagged Dectin-1 or the ELAM-luciferase reporter have been described previously[8]. Dectin-1-deficient mice were provided by G. D. Brown. Culture of primary mouse macrophages and dendritic cells was performed as in previous studies[22]. Human monocytes were obtained from peripheral blood, and macrophages were derived by 7-day culture with 50 ng ml^{-1} rhM-CSF. Cytokine and reactive oxygen species production, Syk phosphorylation, MAP kinase and NF-κB activation, and Egr2/3 induction were assessed as previously described[11,13,22]. Active phospho-Syk (Y525/Y526), active phospho-Src family kinases (Y416) and inactive phospho-Lyn (Y507) antibodies were from Cell Signalling Technology.

β-glucan immobilization and CD45 co-immobilization. Soluble β-glucans were immobilized on tissue culture plates or large polystyrene latex beads (0.8 and 3 μm; Sigma) by incubation with PBS/EDTA containing 100 μg ml^{-1} soluble β-glucan for 1 h at 37 °C. Plates or beads were then washed three times with PBS/EDTA to remove unbound β-glucans, and blocked with media containing 10% FCS for 30 min before use. For CD45 co-immobilization assays, soluble HMW β-glucan (20 μg ml^{-1}) and LPS (100 ng ml^{-1}) were immobilized on tissue culture plates in PBS/EDTA in the presence/absence of 10 ng ml^{-1} anti-CD45 or rat IgG for 1 h at 37 °C, and washed and blocked as above before the addition of BMDC. For assays using small polystyrene latex beads (0.05, 0.2 and 0.5 μm; Polysciences), coating was achieved by incubating beads in PBS/EDTA containing 50 μg ml^{-1} 5-(4,6-dichlorotriazinyl) aminofluorescein (DTAF)-labelled soluble HMW β-glucan, and beads were diluted to a final concentration of 0.05 μg ml^{-1} β-glucan for stimulation, a dose at which the soluble β-glucans are too dilute to block Dectin-1 signalling by β-glucan particles (data not shown). Beads were fed to cells at a dose that achieves presentation of an equivalent total β-glucan-coated surface area per cell (approximately 40:1 0.5 μm beads:cell, 250:1 0.2 μm beads:cell, and 4,000:1 0.05 μm beads:cell).

Soluble β-glucan binding to Dectin-1. Parental RAW264.7 macrophages or RAW264.7 macrophages stably expressing SBPc-tagged-Dectin-1 were incubated for 2 h at 37 °C with 100 μg ml^{-1} unlabelled MMW soluble β-glucan, and binding was detected by flow cytometry using a mouse IgM monoclonal antibody specific for β-(1,3)-linked glucan (BfD IV; clone 10C6; ref. 24) and a FITC-conjugated goat anti-mouse secondary antibody. Dectin-1 expression by the macrophages was assessed by flow cytometry using a FITC-conjugated anti-Dectin-1 antibody (2A11) from Serotec. Macrophages and dendritic cells incubated with DTAF-labelled soluble β-glucans were washed and fixed before analysis by flow cytometry.

Anti-Dectin-1 competition assay. Cells were incubated on ice in media containing 0.4 μg ml^{-1} FITC-anti-Dectin-1 and the indicated concentrations of unlabelled soluble β-glucans for 30 min, and then washed and fixed before assessment of anti-Dectin-1 binding by flow cytometry.

Confocal microscopy of fixed cells. Cells were plated on glass coverslips overnight before the addition of β-glucan particles (unlabelled zymosan or AlexaFluor647-labelled WGP), brief centrifugation to ensure particle contact with the cells, and incubation for 1 min at 37 °C. Cells were then washed with ice-cold PBS to remove unbound particles, fixed with 10% formalin for 20 min, and permeabilized with ice-cold acetone for 30 s. Nonspecific binding was blocked by incubation with TBS + 5% FCS for 10 min. Cells were stained with unconjugated primary antibodies for 1 h as follows. SBPc-tagged Dectin-1: mouse anti-protein C tag (HPC4; Amersham Biosciences); pSyk: rabbit anti-pSyk (Y525/Y526); pSrc: rabbit anti-pSrc (Y416) (Cell Signalling Technology); pTyr: mouse anti-pTyr (Cell Signalling Technology); CD45: rat anti-CD45 (AbD Serotec); and CD148: hamster anti-CD148 (ref. 17). Cells were then washed and incubated with secondary antibodies for 30 min as follows. Dectin-1 and pTyr: AlexaFluor488-conjugated anti-mouse; pSyk and pSrc: FITC-conjugated anti-rabbit; CD45: AlexaFluor568-conjugated anti-rat; and CD148: AlexaFluor568-conjugated anti-hamster (Invitrogen). AlexaFluor647-conjugated cholera toxin (Invitrogen) was used to stain the plasma membrane of unpermeabilized cells. Coverslips were mounted and examined using a Leica TCS SP5 confocal microscope. Image analysis was performed using Leica LAS AF software, ImageJ and ImageSurfer[23], and Volocity (Perkin Elmer).

Confocal microscopy of live cells. RAW264.7 macrophages stably expressing Dectin-1 tagged with green fluorescent protein (GFP) at the carboxy terminus and CD45-tagged with DsRed at the carboxy terminus were plated on chamber slides overnight before stimulation and maintained at 37 °C during confocal imaging. For assessment of cell contact with β-glucan-coated surfaces, chamber slides were incubated with PBS/EDTA (control) or 100 μg ml^{-1} β-glucan in PBS/EDTA for 1 h at 37 °C and washed three times with PBS/EDTA before macrophage addition.

Statistics. Statistical significance was assessed using Student's *t*-test.

24. Lavigne, L. M., Albina, J. E. & Reichner, J. S. Beta-glucan is a fungal determinant for adhesion-dependent human neutrophil functions. *J. Immunol.* **177,** 8667–8675 (2006).

LETTER

TLR signalling augments macrophage bactericidal activity through mitochondrial ROS

A. Phillip West[1], Igor E. Brodsky[1]†, Christoph Rahner[2], Dong Kyun Woo[3], Hediye Erdjument-Bromage[4], Paul Tempst[4], Matthew C. Walsh[5], Yongwon Choi[5], Gerald S. Shadel[3] & Sankar Ghosh[6]

Reactive oxygen species (ROS) are essential components of the innate immune response against intracellular bacteria and it is thought that professional phagocytes generate ROS primarily via the phagosomal NADPH oxidase machinery[1]. However, recent studies have suggested that mitochondrial ROS (mROS) also contribute to mouse macrophage bactericidal activity, although the mechanisms linking innate immune signalling to mitochondria for mROS generation remain unclear[2–4]. Here we demonstrate that engagement of a subset of Toll-like receptors (TLR1, TLR2 and TLR4) results in the recruitment of mitochondria to macrophage phagosomes and augments mROS production. This response involves translocation of a TLR signalling adaptor, tumour necrosis factor receptor-associated factor 6 (TRAF6), to mitochondria, where it engages the protein ECSIT (evolutionarily conserved signalling intermediate in Toll pathways), which is implicated in mitochondrial respiratory chain assembly[5]. Interaction with TRAF6 leads to ECSIT ubiquitination and enrichment at the mitochondrial periphery, resulting in increased mitochondrial and cellular ROS generation. ECSIT- and TRAF6-depleted macrophages have decreased levels of TLR-induced ROS and are significantly impaired in their ability to kill intracellular bacteria. Additionally, reducing macrophage mROS levels by expressing catalase in mitochondria results in defective bacterial killing, confirming the role of mROS in bactericidal activity. These results reveal a novel pathway linking innate immune signalling to mitochondria, implicate mROS as an important component of antibacterial responses and further establish mitochondria as hubs for innate immune signalling.

The phagocytic response of the innate immune system involves the production of ROS via the phagosomal NADPH-oxidase-dependent respiratory burst. This is a necessary effector response for the destruction of intracellular microbes[1,6]. In addition to NADPH-oxidase, the mitochondrial oxidative phosphorylation machinery generates ROS when electrons prematurely escape oxidative phosphorylation complexes I and III and react with molecular oxygen to generate superoxide[7,8]. Mitochondria are major sites of ROS production in most cells; however, mROS have traditionally been regarded as byproducts of oxidative respiration and therefore their synthesis was believed to be unregulated[7,9]. To examine whether TLR signalling could enhance mROS production we stimulated RAW 264.7 macrophages (RAW cells) with lipopolysaccharide (LPS, a TLR4 agonist), synthetic lipopeptide Pam3CSK4 (TLR1 and TLR2 agonist), lipotechoic acid (LTA, TLR2 agonist), poly(I:C) (TLR3 agonist), R848 (TLR7 and TLR8 agonist) and CpG DNA (TLR9 agonist) (Fig. 1a). The production of mROS was triggered only upon signalling from the cell-surface TLRs (TLR1, TLR2 and TLR4), whereas stimulation of endosomal TLRs (TLR3, TLR7, TLR8 and TLR9) failed to augment mROS (Fig. 1a). Exposure of cells to rotenone and antimycin A, compounds

known to increase mitochondrial superoxide generation, did augment mROS but treatment with tumour-necrosis factor-α (TNF-α) did not (Fig. 1a)[7]. We observed similar increases in mROS when bone-marrow-derived macrophages (BMDMs) were stimulated with TLR1, TLR2 and TLR4 (TLR1/2/4) agonists but we were again unable to detect significant induction of mROS upon ligation of TLR9 (Fig. 1b). We also detected increased cellular hydrogen peroxide (H_2O_2) generation upon TLR2/TLR4 ligation, but not after TLR9 ligation (Fig. 1b)[10–12]. As ROS are critical for antibacterial responses,
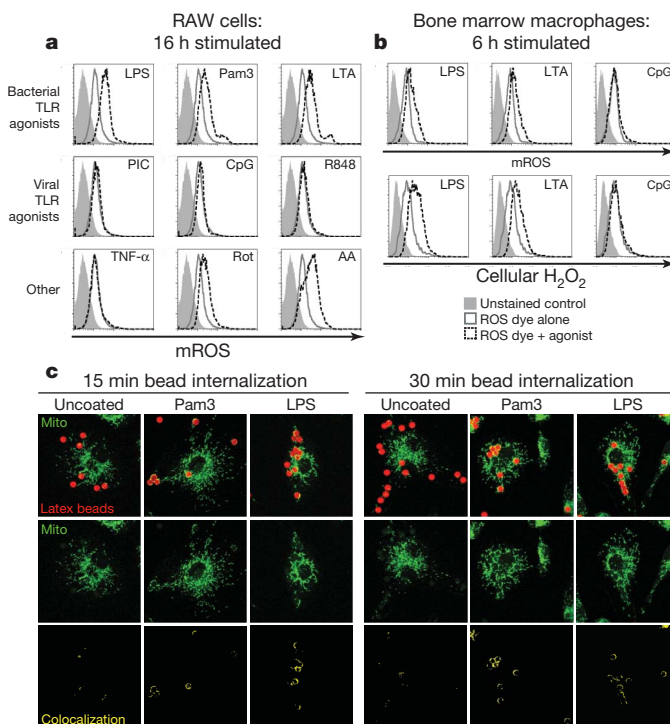

Figure 1 | TLR1/2/4 signalling induces mROS generation and mitochondrial recruitment to phagosomes. a, RAW cells stimulated as indicated, stained with MitoSOX (mROS) and analysed by fluorescence-activated cell sorting (FACS). Pam3, Pam3CSK4; PIC, poly(I:C); Rot, rotenone; AA, antimycin A. **b**, BMDMs stimulated as indicated, stained with MitoSOX (top panels) or CM-H_2DCFDA (cellular H_2O_2, bottom panels) and analysed by FACS. **c**, BMDMs were incubated with uncoated, Pam3CSK4-coated or LPS-coated latex beads, mitochondrial networks were immunostained with HSP70 antibodies (Mito) and confocal Z-stacks were acquired. Co-localized beads (red pixels) and mitochondria (green pixels) are displayed in yellow (bottom). Images shown are representative of approximately 100 cells analysed at ×63 original magnification.

[1]Department of Immunobiology, Yale University School of Medicine, New Haven, Connecticut 06520, USA. [2]Department of Cell Biology, Yale University School of Medicine, New Haven, Connecticut 06520, USA. [3]Department of Pathology, Yale University School of Medicine, New Haven, Connecticut 06520, USA. [4]Molecular Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA. [5]Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104, USA. [6]Department of Microbiology and Immunology, College of Physicians and Surgeons, Columbia University, New York, New York 10032, USA. †Present address: Department of Pathobiology, University of Pennsylvania School of Veterinary Medicine, Philadelphia, Pennsylvania 19104, USA.

it is not surprising that signalling from cell-surface TLRs, which predominantly recognize ligands derived from bacteria, induces ROS generation[13]. In contrast, ROS are not used as direct antiviral effectors and hence endosomal TLRs, which function primarily in sensing viral infection, do not seem to augment ROS production.

Several reports have indicated that mitochondria are recruited to vacuoles containing intracellular pathogens[14–17]. To investigate whether recruitment of mitochondria to phagosomes might be an active process mediated by innate immune signalling, we examined mitochondrial localization in cells loaded with latex beads coated with pathogen-associated molecular patterns (PAMPs). Such coated beads have been used previously to investigate signalling in phagocytic cells and have been shown to recruit innate immune signalling components, analogous to phagocytosed bacteria[18,19]. Interestingly, we observed mitochondrial recruitment and cupping around Pam3CSK4- and LPS-coated beads in BMDMs (Fig. 1c). Uncoated beads, despite being taken up by BMDMs to a similar extent, did not co-localize efficiently with mitochondrial networks and displayed markedly lower mitochondrial cupping per bead (Fig. 1c and Supplementary Fig. 2).

On the basis of the above findings, we hypothesized that the inducible juxtaposition of phagosomes and mitochondria should be accompanied by the concomitant translocation of TLR signalling components. A key intermediate in TLR1/2/4 signalling is TRAF6 and immunoblotting of highly purified cellular extracts from LPS-stimulated macrophages revealed that TRAF6 was enriched in mitochondrial fractions (Fig. 2a). This recruitment was specific to TRAF6 because other cytosolic proteins that interact transiently with TLR signalling complexes, such as MyD88, IRAK4 (not shown), IRAK1, TAK1, and IκBα, were not detected in mitochondrial fractions. Furthermore, stimulation with Pam3CSK4 and LTA induced TRAF6 recruitment to mitochondria with similar kinetics to that triggered by LPS (Fig. 2b). Consistent with the results concerning mROS generation, we were unable to detect TRAF6 in the mitochondrial fractions of macrophages stimulated with poly(I:C) or CpG (Fig. 2c).

The induction of mROS and the recruitment of TRAF6 to mitochondria upon TLR1/2/4 stimulation indicated that TRAF6 might interact with mitochondrial proteins to control mROS production. Recent studies have shown that ECSIT, a previously characterized TRAF6-interacting protein, localizes to mitochondria and has a role in oxidative phosphorylation complex I assembly[5,20,21]. Mass spectrometry analysis of purified ECSIT protein complexes confirmed that ECSIT associates with oxidative phosphorylation complex I components (Supplementary Table 1). Immunofluorescence microscopy and biochemical fractionation experiments revealed that ECSIT localizes

predominantly to mitochondria in both fibroblasts and BMDMs (Supplementary Fig. 3a–c). Additional analysis confirmed that ECSIT localizes to the inner mitochondrial membrane, consistent with its role in complex I assembly[5]. However, we also observed some ECSIT molecules proximal to outer mitochondrial membranes (OMMs), indicating that OMM-associated ECSIT might interact with TRAF6 recruited from phagosomal TLR signalling complexes (Supplementary Fig. 3d–f). Accordingly, we detected inducible interactions between ECSIT and TRAF6 in purified mitochondrial extracts from macrophages stimulated with LPS (Fig. 2d).

TRAF6 possesses E3-ubiquitin ligase activity; therefore, we explored whether ECSIT is ubiquitinated by TRAF6 (ref 22, 23). ECSIT was polyubiquitinated when co-transfected with TRAF6 in HEK 293 cells (Supplementary Fig. 4a) and a dominant-negative form of ECSIT lacking the TRAF6 interaction domain was significantly less ubiquitinated by TRAF6 than wild-type ECSIT (Supplementary Fig. 4b)[21]. We also detected an increase in ECSIT polyubiquitination in macrophages after exposure to LPS, which mirrored the mitochondrial recruitment kinetics of TRAF6 (Supplementary Fig. 4c). In addition, total LPS-induced ECSIT ubiquitination was decreased in TRAF6-knockdown macrophages, indicating a requirement for TRAF6 in the ubiquitination of ECSIT during TLR4 signalling (Supplementary Fig. 4c).

We next investigated the dynamics of ECSIT localization in mitochondria after LPS stimulation. Notably, after 30 min of LPS treatment, ECSIT became more sensitive to proteinase K in the absence of OMM permeabilization by saponin (Fig. 2e). In contrast, the inner mitochondrial membrane protein NDUFS3 remained largely insensitive to proteinase K without saponin treatment. This indicates that ECSIT becomes enriched at the mitochondrial periphery, and thus becomes more sensitive to protease digestion, upon LPS signalling. Electron microscopy analysis further confirmed these data, as more ECSIT was localized peripheral to the OMM after LPS treatment (Supplementary Fig. 5). Protease sensitivity assays on mitochondria from TRAF6-knockdown RAW cells indicated that TRAF6 is required for LPS-induced ECSIT enrichment on OMMs (Supplementary Fig. 6, compare lanes 5 and 7 with 9 and 11). In addition to influencing the mitochondrial localization of ECSIT, TRAF6–ECSIT signalling also appears to regulate the recruitment of mitochondria around PAMP-coated latex beads. Both TRAF6-knockout and ECSIT-knockdown BMDMs showed less mitochondrial enrichment around phagosomes containing LPS- and Pam3CSK4-coated beads than did wild-type BMDMs (Supplementary Fig. 7).

The link between ECSIT and complex I led us to hypothesize that ECSIT might modulate mROS derived from this complex[5,7]. To
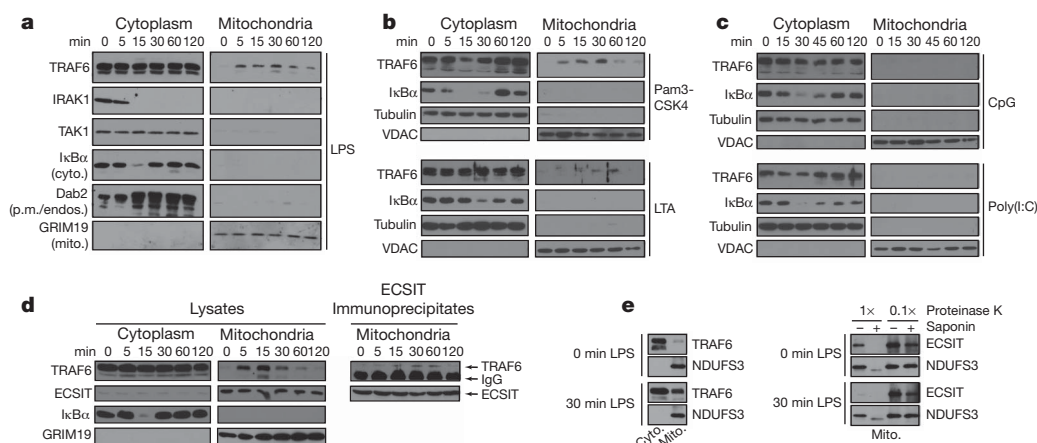

Figure 2 | TRAF6 is recruited to mitochondria upon TLR1/2/4, but not TLR3/9, signalling to engage ECSIT on the mitochondrial surface.
a–e, RAW cells were stimulated with TLR agonists for the indicated times, cells were fractionated and extracts were western-blotted. cyt., cytoplasmic localization; p.m./endos., plasma membrane/endosomes localization; mito.,

mitochondrial localization. **d**, Purified mitochondrial lysates were immunoprecipitated with ECSIT antibody overnight. **e**, Equal amounts of extracts were treated with the indicated amount of proteinase K ($1\times$, $33\,ng\,\mu l^{-1}$; $0.1\times$, $3.3\,ng\,\mu l^{-1}$) on ice with or without 0.2% saponin to permeabilize mitochondrial membranes.
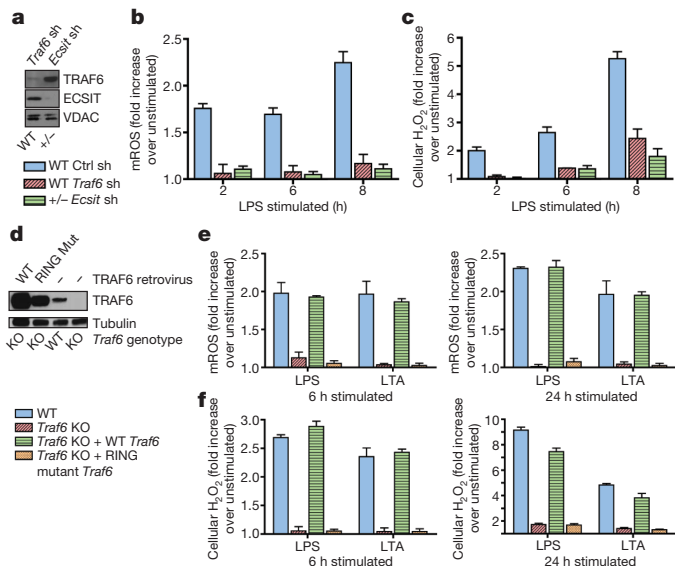
Figure 3 | TRAF6-ECSIT signalling regulates the generation of mitochondrial and cellular ROS, which requires TRAF6 E3-ubiquitin ligase activity. a–c, Wild-type (WT) or $Ecsit^{+/-}$ (+/−) BMDMs were transduced with short hairpin RNAs (sh), then either left untreated or stimulated with LPS. a, Untreated BMDMs were lysed and extracts blotted for TRAF6 and ECSIT. b, c, Cells were stained with MitoSOX (b) or CM-H$_2$DCFDA (c) and analysed by FACS. d–f, wild-type or TRAF6-null BMDMs were left alone or transduced with TRAF6-expressing retroviruses, then either left unstimulated or stimulated for the indicated times with LPS or LTA. d, Unstimulated BMDMs were lysed and extracts blotted for TRAF6 expression. e, f, Cells were stained with MitoSOX (e) or CM-H$_2$DCFDA (f) and analysed by FACS. All error bars represent s.d. of the mean from triplicate samples.

embryonic lethals[24]. Heterozygous $Ecsit^{+/-}$ animals are viable but have ~40% less ECSIT than wild-type animals (Supplementary Fig. 8a). Interestingly, BMDMs from $Ecsit^{+/-}$ mice generated modestly lower mROS and cellular H$_2$O$_2$ when stimulated with LPS and LTA (Supplementary Fig. 8b, c). To confirm the importance of TLR1/2/4-induced TRAF6–ECSIT signalling in mROS responses, we analysed $Traf6$- and $Ecsit$-knockdown BMDMs (Fig. 3a and Supplementary Fig. 11a). Upon LPS (Fig. 3b) or LTA (Supplementary Fig. 9) stimulation, we observed marked reductions in mROS production in both ECSIT- and TRAF6-depleted macrophages at all time points tested. Although TLR-generated mROS responses were impaired in $Ecsit$-knockdown cells, mROS production elicited by rotenone or antimycin
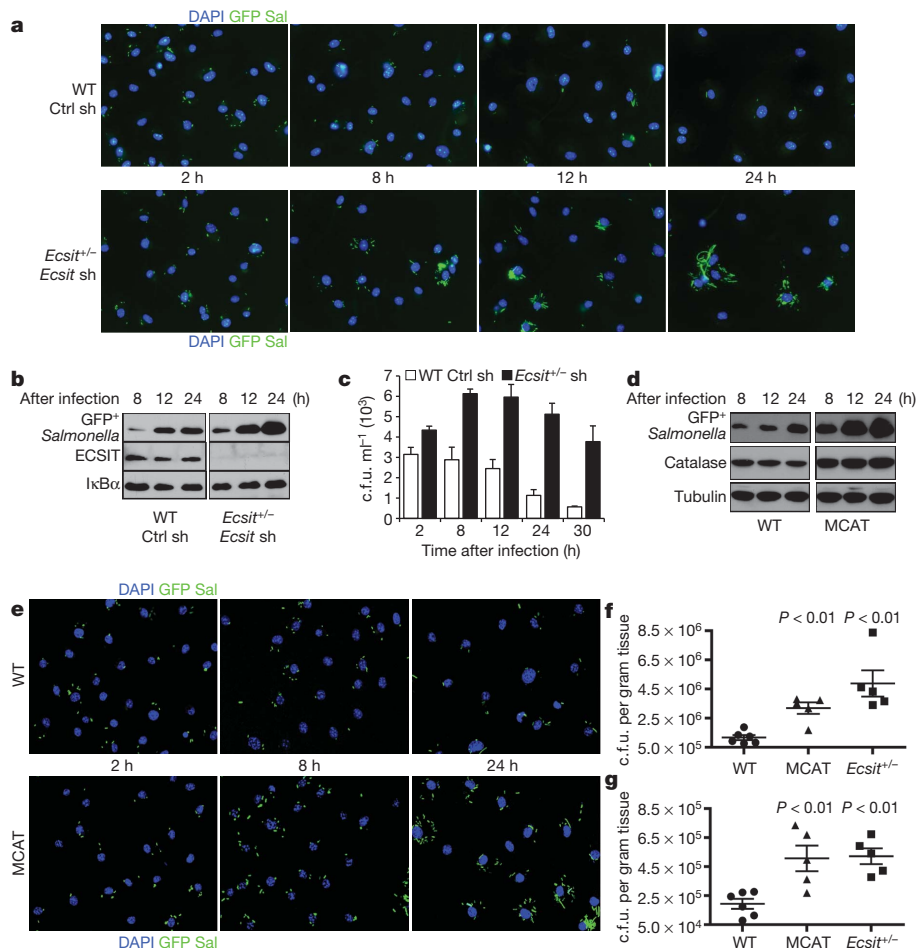
establish the role of ECSIT in TLR1/2/4-dependent upregulation of mROS, we sought to examine macrophages lacking ECSIT. Although $Ecsit$-knockout mice have been generated, they are very early

Figure 4 | ECSIT-depleted and MCAT transgenic macrophages are less effective at clearing *Salmonella* than wild-type macrophages. a–c, BMDMs from wild-type or $Ecsit^{+/-}$ mice were transduced with shRNAs and infected with GFP-expressing *Salmonella* (GFP Sal). Cells were fixed and stained with 4′,6-diamidino-2-phenylindole (DAPI) (a), solubilized in SDS (b) or lysed and plated (c). d, e, wild-type or MCAT BMDMs were infected with GFP-expressing *Salmonella*. Cells were solubilized in SDS (d) or fixed and DAPI-stained (e). Triplicate wells were pooled and blotted (b, d), and error bars (c) represent s.d. from triplicate samples. f, g, wild-type ($n = 6$), MCAT ($n = 5$) and $Ecsit^{+/-}$ ($n = 5$) mice were infected with *Salmonella* intraperitoneally. Five days after infection, spleens (f) and livers (g) were homogenized and colony-forming units (c.f.u.) per gram of tissue were determined. Error bars indicate s.e.m. and *P* values are relative to wild-type.

A was unaffected (Supplementary Fig. 10). In addition, LPS- (Fig. 3c and Supplementary Fig. 11b), LTA- (Supplementary Fig. 9) and Pam3CSK4-induced (Supplementary Fig. 11c) cellular H_2O_2 levels were markedly reduced upon *Ecsit* and *Traf6* knockdown. Thus, induction of mROS and cellular H_2O_2 by bacterial PAMPs is critically dependent on both TRAF6 and ECSIT. To determine whether the E3-ubiquitin ligase activity of TRAF6 is required for ROS generation, we examined *Traf6*-knockout BMDMs reconstituted with either wild-type or RING (really interesting new gene) finger domain mutant *Traf6* constructs (Fig. 3d)[25]. In agreement with the knockdown results, TRAF6-null BMDMs generated markedly lower levels of mROS and cellular H_2O_2 in response to both LPS and LTA (Fig. 3e, f). Null macrophages reconstituted with wild-type, but not RING mutant, TRAF6 regained the ability to generate ROS in response to TLR2/4 agonists (Fig. 3e, f). Therefore, these data indicate that a functional RING domain is required for TRAF6 signalling to induce ROS generation, probably because TRAF6 mediates ECSIT ubiquitination (Supplementary Fig. 4).

To test the functional significance of these findings, we assessed the responses of ECSIT- and TRAF6-deficient macrophages to *Salmonella typhimurium*, a Gram-negative, facultative intracellular pathogen that is sensitive to ROS-dependent killing[26–28]. Consistent with the results obtained using purified PAMPs, we observed decreased mitochondrial and cellular ROS in *Ecsit*- and *Traf6*-knockdown BMDMs when exposed to *S. typhimurium* (Supplementary Fig. 12). To determine whether mROS are important in macrophage bactericidal responses, control or *Ecsit*-knockdown BMDMs were infected with GFP-expressing *S. typhimurium* and analysed by immunofluorescence microscopy and western blotting. Notably, ECSIT-depleted BMDMs harboured significantly more GFP-expressing *S. typhimurium* than control cells (Fig. 4a, b). The direct measurement of intracellular bacterial colony forming units showed that there were significantly more bacteria in ECSIT-deficient cells than in control cells at all time points examined (Fig. 4c). The reduced ability of ECSIT-deficient BMDMs to control intracellular bacteria was not the result of a non specific impairment of NADPH-oxidase activity, as the PMA-stimulated respiratory burst was unaffected in *Ecsit* knockdowns (Supplementary Fig. 13). Additionally, nitric oxide and proinflammatory cytokine production were similar in control and knockdown BMDMs, collectively indicating that ECSIT depletion does not result in systemic innate immune deficiency (Supplementary Fig. 14).

Mitochondria are regarded as a significant source of H_2O_2 in most cell types and peroxisomal catalase converts H_2O_2 into water and oxygen, reducing oxidative damage caused by these ROS[7]. Over-expressing catalase in mitochondria using transgenic approaches (MCAT mice) leads to significantly lower mitochondrial H_2O_2 levels and these transgenic mice show lower levels of age-related oxidative damage and an extended lifespan[29]. Consistent with our finding that mROS contribute to total cellular ROS levels, MCAT BMDMs generated significantly less LPS-induced cellular H_2O_2 than wild-type cells (Supplementary Fig. 15). To test the specific role of H_2O_2 derived from mitochondria in controlling intracellular bacterial replication, wild-type or MCAT BMDMs were challenged with GFP-expressing *S. typhimurium*. Similar to *Ecsit* knockdowns, MCAT BMDMs had significantly higher bacterial loads than wild-type cells between 8 and 24 h after infection (Fig. 4d, e). Finally, to confirm the role of mROS in control of bacterial infection *in vivo*, we challenged wild-type, MCAT and *Ecsit*[+/−] mice by intraperitoneal infection with *S. typhimurium* and measured bacterial burdens in the spleen and liver five days after infection. In agreement with data from isolated BMDMs, both MCAT and *Ecsit*[+/−] mice harboured roughly twofold to three fold more bacteria per gram of tissue than wild-type littermates, further substantiating the notion that mROS play an integral part in antibacterial innate immunity (Fig. 4f, g).

We have discovered a novel pathway by which macrophages generate ROS in response to bacteria by coupling TLR1/2/4 signalling to mitochondrial complex I via TRAF6 and ECSIT (Supplementary Fig. 1). Our study demonstrates that in addition to NADPH-oxidase-derived ROS, mROS have an important role in macrophage innate immunity. To our knowledge, this is the first evidence of direct communication between TLRs and mitochondria. This study also highlights a remarkable symmetry between mitochondrial antiviral signalling protein (MAVS) and ECSIT in innate immune responses. As the clearance of intracellular bacteria requires ROS, TRAF6-ECSIT signalling is engaged downstream of bacterial PAMP-sensing TLRs for robust ROS production; likewise, MAVS signalling is activated by virus-sensing RIG-I-like (also known as DDX58) receptors to induce type I interferon production and effective antiviral immunity. Our current findings further solidify the emerging idea that mitochondria serve as hubs for innate immune signalling and the generation of effector responses.

METHODS SUMMARY

Animal strains. *Ecsit*-heterozygous and MCAT mice were previously described and maintained on a C57BL/6 background[24,29].

ROS measurements. Cells were incubated with MitoSOX (to measure the mROS superoxide) and/or CM-H_2DCFDA (to measure total cellular H_2O_2) (Invitrogen) at 2.5 μM final concentration. Unstained controls were treated similarly, except that treatments and dyes were omitted.

Cellular fractionation and mitochondrial isolation. Cell pellets were swelled on ice for 10 min, homogenized with a motorized Teflon pestle, then 2.5× MS buffer (525 mM mannitol, 175 mM sucrose, 12.5 mM Tris-HCl, pH 7.5, 12.5 mM EDTA) was added to 1× final buffer concentration. The homogenate was centrifuged three times at 980g for 10 min to pellet nuclei. The supernatant was transferred to a fresh tube and spun at 17,000g for 30 min to pellet mitochondria and the resulting cytosolic supernatant was saved for further analysis. The mitochondrial pellet was washed three times with 1× MS buffer, re-suspended in Sucrose-TE (20% sucrose, 50 mM Tris-HCl, pH 7.5, 10 mM EDTA) and then purified additionally by centrifugation through a 1.0–1.5 M sucrose step gradient at 100,000g for 60 min. Highly purified heavy mitochondria were removed from the interface of the two sucrose solutions by pipetting and transferred to a fresh tube. Mitochondria were then washed three times with 1 ml of 1× MS buffer and pelleted by centrifugation at 13,000 r.p.m. for 10 min.

Bacterial challenge. Gentamicin protection assays were performed using *S. typhimurium* SL1344 transformed with a GFP expression plasmid essentially as described[30]. *In vivo* challenges were performed by intraperitoneally injecting *S. typhimurium* SL1344 bacteria in 100 μl PBS. Five days after infection, mice were killed, livers and spleens were isolated and tissues were weighed and homogenized in PBS.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Lambeth, J. D. NOX enzymes and the biology of reactive oxygen. *Nature Rev. Immunol.* **4,** 181–189 (2004).
2. Arsenijevic, D. *et al.* Disruption of the uncoupling protein-2 gene in mice reveals a role in immunity and reactive oxygen species production. *Nature Genet.* **26,** 435–439 (2000).
3. Rousset, S. *et al.* The uncoupling protein 2 modulates the cytokine balance in innate immunity. *Cytokine* **35,** 135–142 (2006).
4. Sonoda, J. *et al.* Nuclear receptor ERRα and coactivator PGC-1β are effectors of IFN-γ-induced host defense. *Genes Dev.* **21,** 1909–1920 (2007).
5. Vogel, R. O. *et al.* Cytosolic signaling protein Ecsit also localizes to mitochondria where it interacts with chaperone NDUFAF1 and functions in complex I assembly. *Genes Dev.* **21,** 615–624 (2007).
6. Underhill, D. M. & Ozinsky, A. Phagocytosis of microbes: complexity in action. *Annu. Rev. Immunol.* **20,** 825–852 (2002).
7. Koopman, W. *et al.* Mammalian mitochondrial complex I: Biogenesis, regulation and reactive oxygen species generation. *Antioxid. Redox Signal.* 10.1089/ars.2009.2743 (2009).
8. Murphy, M. P. How mitochondria produce reactive oxygen species. *Biochem. J.* **417,** 1–13 (2009).
9. Arnoult, D., Carneiro, L., Tattoli, I. & Girardin, S. E. The role of mitochondria in cellular defense against microbial infection. *Semin. Immunol.* **21,** 223–232 (2009).
10. Adachi, Y. *et al.* IFN-γ primes RAW264 macrophages and human monocytes for enhanced oxidant production in response to CpG DNA via metabolic signaling: roles of TLR9 and myeloperoxidase trafficking. *J. Immunol.* **176,** 5033–5040 (2006).

11. Remer, K. A., Reimer, T., Brcic, M. & Jungi, T. W. Evidence for involvement of peptidoglycan in the triggering of an oxidative burst by *Listeria monocytogenes* in phagocytes. *Clin. Exp. Immunol.* **140,** 73–80 (2005).
12. Werling, D., Hope, J. C., Howard, C. J. & Jungi, T. W. Differential production of cytokines, reactive oxygen and nitrogen by bovine macrophages and dendritic cells stimulated with Toll-like receptor agonists. *Immunology* **111,** 41–52 (2004).
13. West, A. P., Koblansky, A. A. & Ghosh, S. Recognition and signaling by toll-like receptors. *Annu. Rev. Cell Dev. Biol.* **22,** 409–437 (2006).
14. Chong, A., Lima, C. A., Allan, D. S., Nasrallah, G. K. & Garduño, R. A. The purified and recombinant *Legionella pneumophila* chaperonin alters mitochondrial trafficking and microfilament organization. *Infect. Immun.* **77,** 4724–4739 (2009).
15. Horwitz, M. A. Formation of a novel phagosome by the Legionnaires' disease bacterium (*Legionella pneumophila*) in human monocytes. *J. Exp. Med.* **158,** 1319–1331 (1983).
16. Matsumoto, A., Bessho, H., Uehira, K. & Suda, T. Morphological studies of the association of mitochondria with chlamydial inclusions and the fusion of chlamydial inclusions. *J. Electron Microsc. (Tokyo)* **40,** 356–363 (1991).
17. Sinai, A. P., Webster, P. & Joiner, K. A. Association of host cell endoplasmic reticulum and mitochondria with the *Toxoplasma gondii* parasitophorous vacuole membrane: a high affinity interaction. *J. Cell Sci.* **110,** 2117–2128 (1997).
18. Blander, J. M. & Medzhitov, R. Regulation of phagosome maturation by signals from toll-like receptors. *Science* **304,** 1014–1018 (2004).
19. Blander, J. M. & Medzhitov, R. On regulation of phagosome maturation and antigen presentation. *Nature Immunol.* **7,** 1029–1035 (2006).
20. Calvo, S. *et al.* Systematic identification of human mitochondrial disease genes through integrative genomics. *Nature Genet.* **38,** 576–582 (2006).
21. Kopp, E. *et al.* ECSIT is an evolutionarily conserved intermediate in the Toll/IL-1 signal transduction pathway. *Genes Dev.* **13,** 2059–2071 (1999).
22. Chen, Z. J. & Sun, L. J. Nonproteolytic functions of ubiquitin in cell signaling. *Mol. Cell* **33,** 275–286 (2009).
23. Bhoj, V. G. & Chen, Z. J. Ubiquitylation in innate and adaptive immunity. *Nature* **458,** 430–437 (2009).
24. Xiao, C. *et al.* Ecsit is required for Bmp signaling and mesoderm formation during mouse embryogenesis. *Genes Dev.* **17,** 2933–2949 (2003).
25. Deng, L. *et al.* Activation of the IkappaB kinase complex by TRAF6 requires a dimeric ubiquitin-conjugating enzyme complex and a unique polyubiquitin chain. *Cell* **103,** 351–361 (2000).
26. Shiloh, M. U. *et al.* Phenotype of mice and macrophages deficient in both phagocyte oxidase and inducible nitric oxide synthase. *Immunity* **10,** 29–38 (1999).
27. Vazquez-Torres, A. & Fang, F. C. Oxygen-dependent anti-*Salmonella* activity of macrophages. *Trends Microbiol.* **9,** 29–33 (2001).
28. Vazquez-Torres, A. & Fang, F. C. *Salmonella* evasion of the NADPH phagocyte oxidase. *Microbes Infect.* **3,** 1313–1320 (2001).
29. Schriner, S. E. *et al.* Extension of murine life span by overexpression of catalase targeted to mitochondria. *Science* **308,** 1909–1911 (2005).
30. Brodsky, I. E., Ghori, N., Falkow, S. & Monack, D. Mig-14 is an inner membrane-associated protein that promotes Salmonella typhimurium resistance to CRAMP, survival within activated macrophages and persistent infection. *Mol. Microbiol.* **55,** 954–972 (2005).

METHODS

Animal strains. *Ecsit* heterozygous and MCAT mice were previously described and maintained on a C57BL/6 background[24,29].

Cell lines and reagents. RAW 264.7, J774A.1 and NOR10 cells were obtained from the ATCC and maintained in DMEM (Invitrogen) containing 5% fetal bovine serum (FBS) (Atlanta Biological). *Escherichia coli* LPS (L6529) and all other chemicals were obtained from Sigma unless otherwise noted. TNF-α was from R&D Systems. CpG 1826 oligonucleotides were synthesized by IDT. *Staphylococcus aureus* LTA, R848 and Pam3CSK4 were purchased from Invivogen. Rabbit polyclonal antibodies against ECSIT were previously described[24]. The following antibodies were also used: mouse monoclonal NDUFS3, VDAC, GRIM19, cytochrome *c* (Mitosciences); goat polyclonal HSP60 and HSP70, rabbit polyclonal IRAK1 and IκBα, mouse monoclonal ubiquitin (Santa Cruz Biotechnology); mouse monoclonal COX1 (Molecular Probes); mouse monoclonal catalase, β-tubulin, Flag M2, HA (Sigma); rabbit monoclonal TRAF6 (Abcam); rabbit polyclonal MAVS and TAK1 (Cell Signaling Technology); mouse monoclonal Dab2 and GFP (BD Biosciences).

ROS measurements. Cells (RAW or BMDMS) were plated in non-tissue-culture-treated six-well dishes and treated with various agonists or stimulated with TLR ligands as indicated. Concentrations of stimulants were as follows: 500 ng ml^{-1} LPS, 1 μg ml^{-1} Pam3CSK4, 2 μg ml^{-1} LTA, 10 μg ml^{-1} poly(I:C), 1 μM CpG, 1 μg ml^{-1} R848, 10 ng ml^{-1} TNF-α, 500 nM rotenone, or 5 μM antimycin A. Culture medium was removed, cells were washed with PBS, then incubated with MitoSOX (to measure the mROS superoxide) and/or CM-H$_2$DCFDA (to measure total cellular H$_2$O$_2$) (Invitrogen) at 2.5 μM final concentration in serum-free DMEM (Invitrogen) for 15 to 30 min at 37 °C. Cells were washed with warmed PBS (37 °C), removed from plates with cold PBS containing 1 mM EDTA by pipetting, pelleted at 1,500 r.p.m. for 3 min, immediately re-suspended in cold PBS containing 1% FBS and subjected to fluorescence-activated cell sorting (FACS) analysis. Unstained controls were treated similarly, except that treatments and dyes were omitted. To control for baseline dye fluorescence, samples from each experiment were left unstimulated but stained according to the above procedure. FACS mean fluorescence intensity values (Figs 3b, c, e, f and Supplementary Figs 11b, c and Fig. 15) were calculated by dividing TLR-stimulated by unstimulated values. Error bars were generated by calculating standard deviation (s.d.) of the mean from triplicate samples. All ROS experiments shown are representative of three independent experiments.

Latex bead phagocytosis assays. Yellow orange 3 micron Fluoresbrite carboxy latex microparticles (Polysciences) were left untreated or coated with 30 μg ml^{-1} LPS or Pam3CSK4 overnight at 4 °C in PBS. Beads were washed ten times in large volumes of PBS containing 1% FBS to remove unbound TLR agonists. BMDMs were cultured on coverslips in 12-well dishes and incubated on ice for 10 min. Microparticles were added at a concentration of approximately 4–8 beads per cell and allowed to settle on the cells for an additional 10 min on ice. Warm medium was added to the cells and plates were moved to 37 °C to allow bead phagocytosis for the indicated times. Cells were then fixed for 20 min at room temperature with paraformaldehyde, permeabilized with 0.1% Triton X-100 and stained with the indicated primary and appropriately labelled secondary antibodies for confocal microscopy. Co-localization and quantification analysis was performed using ImageJ software. Three large fields of view were collected for each condition (each field of view contained roughly 30 cells and approximately 120 yellow orange beads) and the integrated density of the colocalized beads (red pixels) and mitochondria (green pixels) was determined for each image. This value was divided by the total number of beads in each field to determine colocalized pixels per bead and the mean and s.d. were calculated by averaging the three fields of view for each condition. The percentage of beads with mitochondrial cupping (defined as colocalized mitochondria surrounding ~20% or more of the bead periphery) was determined by analysing the same fields of view described above and calculating the mean and s.d. of the percentage values obtained.

Cellular fractionation and mitochondrial isolation. PBS-washed cell pellets were re-suspended in ten pellet volumes of RSB buffer (10 mM NaCl, 1.5 mM CaCl$_2$, 10 mM Tris-HCl, pH 7.5), swelled on ice for 10 min, homogenized with a motorized Teflon pestle, then 2.5× MS buffer (525 mM mannitol, 175 mM sucrose, 12.5 mM Tris-HCl, pH 7.5, 12.5 mM EDTA) was added to 1× final buffer concentration. The homogenate was centrifuged three times at 980*g* for 10 min to pellet nuclei, which were saved for western blot analysis. The supernatant was transferred to a fresh tube and spun at 17,000*g* for 30 min to pellet mitochondria and the resulting cytosolic supernatant was saved for further analysis. The mitochondrial pellet was washed three times with 1× MS buffer, re-suspended in Sucrose-TE (20% sucrose, 50 mM Tris-HCl, pH 7.5, 10 mM EDTA), then purified additionally by centrifugation through a 1.0–1.5 M sucrose step gradient at 100,000*g* for 60 min. Highly purified heavy mitochondria were removed from the interface of the two sucrose solutions by pipetting and were then transferred to a fresh tube. Mitochondria were washed three times with 1 ml of 1× MS buffer

and pelleted by centrifugation at 13,000 r.p.m. for 10 min. SDS was then added to nuclear and cytosolic fractions to a final concentration of 1% and mitochondrial pellets were re-suspended in SDS lysis buffer (20 mM Tris-HCl, 1% SDS, pH 7.5) and boiled for 5 min before normalizing protein concentration and western blotting. Fraction purity was confirmed by blotting for IκBα and tubulin (cytoplasm, cyto.), Dab2 (plasma membrane and endosomes, p.m./endos.), and GRIM19 and VDAC (mitochondria, mito.).

Mitochondrial sub-fractionation and Triton X-100 extraction. RAW cell mitochondria were highly purified as above and sub-fractionated as described[31]. For Triton X-100 solubility assay, equal amounts of RAW mitochondria were re-suspended in 1× MS buffer containing the indicated concentrations of Triton X-100 on ice for 20 min. The samples were then spun at 8,000*g* to separate unsolubilized mitochondria from extracted proteins in the supernatant and the resulting pellets were completely solubilized in 1% Triton X-100. Triton concentrations between 0.01 and 0.05% solubilize outer mitochondrial membrane and intermembrane space proteins, whereas concentrations above 0.1% solubilize both outer and inner membrane proteins. Samples were then blotted accordingly.

TRAF6 recruitment and protease sensitivity assays. For the TRAF6 recruitment assay, RAW cells were left untreated or stimulated with TLR agonists (1 μg ml^{-1} LPS, 1 μg ml^{-1} Pam3CSK4, 2 μg ml^{-1} LTA, 1 μM CpG, or 10 μg ml^{-1} poly(I:C)), subjected to mitochondrial isolation as above and fractions were blotted as indicated. For mito-immunoprecipitation, mitochondria were lysed in TNT buffer (25 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1% Triton X-100, 10% glycerol, 1 mM EDTA, protease/phosphatase inhibitors) and incubated with ECSIT antibody and anti-rabbit IgG agarose overnight at 4 °C. The samples were washed extensively and blotted. For the proteinase K sensitivity assay, RAW cells treated with a control shRNA to GFP or with *Traf6* shRNA were left untreated or stimulated with LPS, subjected to mitochondrial fractionation, then purified mitochondria were re-suspended in digest buffer (25 mM Tris-HCl, pH 7.5, 125 mM sucrose, 1 mM CaCl$_2$). Samples were incubated with proteinase K (1×, 33 ng μl^{-1}; 0.1×, 3.3 ng μl^{-1}; ++, 12 ng μl^{-1}; +, 6 ng μl^{-1}) with or without 0.1% saponin on ice for 30 min. Saponin was used to permeabilize mitochondrial membranes gently and allow proteinase K to enter mitochondria. Proteinase K activity was quenched with phenylmethylsulphonyl fluoride (PMSF), SDS was added to 1% to solubilize mitochondrial proteins and samples were blotted as indicated.

Ubiquitination assays. For endogenous ubiquitination assays, one 10-cm dish of J774 control *GFP* shRNA or *Traf6* shRNA cells per time point was stimulated as described and solubilized in boiling SDS lysis buffer, then boiled for an additional 10 min. After shearing DNA, the samples were diluted in 2× TNT buffer to 1× and immunoprecipitated overnight with ECSIT antibody and anti-rabbit IgG beads (eBioscience). Beads were washed in 1× TNT buffer and samples were blotted and probed with antibodies as indicated. All other ubiquitination assays were performed using 6-cm dishes of 293T cells transfected with 0.25–1 μg each of the indicated constructs. After overnight incubation (and 2 h incubation in MG132, if indicated), cells were lysed as described above, diluted in 2× TNT buffer and immunoprecipitated with anti-Flag resin (Sigma) overnight. Beads were washed extensively and samples blotted and probed as indicated. TRAF6, ECSIT-FLAG, and ΔN-ECSIT–Flag were previously described[21].

Bacterial infection and gentamicin protection assay. Gentamicin protection assays were performed using *S. typhimurium* SL1344[32] transformed with a GFP expression plasmid essentially as described[30,33]. Briefly, BMDMs were plated at 2×10^5 cells per well in 24-well dishes, incubated with opsonized (using fresh mouse serum) GFP-expressing *Salmonella* at a multiplicity of infection of 20 for 1 h, then further incubated with 100 μg ml^{-1} gentamicin for 1 h. Wells were washed several times and 2 h time points were collected or cells further incubated in media containing 10 μg ml^{-1} gentamicin. Similar results were obtained when the initial 100 μg ml^{-1} gentamicin incubation was omitted and cells were only exposed to 10 μg ml^{-1} gentamicin (not shown). For colony count analysis, triplicate samples were lysed in PBS containing 1% Triton X-100, serially diluted in PBS and plated on LB agar plates containing streptomycin. For western blot analysis of GFP-expressing *Salmonella*, triplicate BMDMs samples were infected as above and cells were lysed at indicated time points in SDS lysis buffer. After boiling and shearing DNA, samples were pooled and blotted as described. For immunofluorescence analysis of GFP-expressing *Salmonella*, 2×10^5 BMDMs were grown on coverslips in 12-well dishes and treated as above. At indicated time points, cells were fixed with 4% paraformaldehyde for 20 min at room temperature, nuclei stained with DAPI and cells mounted and imaged by confocal microscopy. Experiments shown are representative of at least three independent experiments.

***In vivo Salmonella* infection assay.** 8–10-week-old wild-type, MCAT, or *Ecsit*$^{+/-}$ littermates were injected intraperitoneally with 200 *S. typhimurium* SL1344 bacteria in 100 μl PBS. Five days after infection, mice were killed, livers and spleens were isolated and tissues were weighed and homogenized in 1 ml PBS.

Lysates were serially diluted in PBS and 100 µl of each dilution was plated on LB streptomycin plates. Colonies were counted the next morning with a BioRad VersaDoc imager using the colony counting feature of the Quantity One software. Error bars represent standard error of the mean (s.e.m.).

Tandem affinity purification and mass spectrometry. 293T or RAW cells were transiently or stably transfected with pCTAP-ECSIT (primers used: F, tacgagtcc gaattcccaccatgagctgggtgcaggtcaac; R, tacgagtccctcgagactttgcccctgctgctgctc) respectively and TAP complexes were purified using the Interplay Mammalian TAP System (Stratagene) as described. Purified complexes were precipitated in 10% trichloroacetic acid, washed with cold acetone, air dried and re-suspended in 4× Laemmli sample buffer. Samples were electrophoresed on SDS–polyacrylamide gel electrophoresis (SDS–PAGE) gels until just entering the resolving gel, slices were removed and subjected to in-gel trypsin digestion and mass spectrometry analysis in the laboratory of P. Tempst at Memorial Sloan-Kettering Cancer Center as described[34,35].

Immuno-electron microscopy. J774 cells were left untreated or stimulated with LPS, washed with PBS and fixed (2% paraformaldehyde, 0.1% gluteraldehyde, 3% sucrose, 0.25 M HEPES, pH 7.4) for 30 min at 4 °C. Samples were then rinsed in PBS and re-suspended in 10% gelatin, chilled and trimmed to smaller blocks, then placed in cryoprotectant (2.3 M sucrose) rotating overnight at 4 °C. Small pieces were transferred to aluminium pins and frozen rapidly in liquid nitrogen. The frozen block was trimmed on a Leica Cryo-EMUC6 UltraCut and 75-nm-thick sections were collected using the Tokoyasu method (1973), placed on a nickel formvar/carbon-coated grid and floated in a dish of PBS ready for immunolabelling. Grids were placed section-side-down on drops of 0.1 M ammonium chloride for 10 min to quench untreated aldehyde groups, then blocked for nonspecific binding on 1% fish skin gelatin in PBS for 20 min. Single labelled grids were incubated in primary rabbit anti-ECSIT antibody at 1:20 dilution for 30 min. After rinsing, the grids were placed on protein A gold (UtrechtUMC) for 30 min. All grids were rinsed in PBS, fixed with 1% gluteraldehyde for 5 min, rinsed in water and transferred to a 0.5% uranyl acetate/1.8% methylcellulose drop for 10 min, then collected and dried. Grids were viewed with a FEI Tencai Biotwin TEM at 80 KV. Images were taken using a Morada CCD and item software (Olympus).

Bone marrow macrophage generation and lentiviral shRNA/retroviral transduction. Bone marrow was collected from littermate wild-type, *Ecsit*$^{+/-}$ or MCAT mice and cultured on Petri plates for a period of 7 days in DMEM containing 10% FBS plus 30% L929 conditioned media. Media was replenished on day four of culture. Lentiviral shRNA constructs were purchased from Open Biosystems (pLKO.1 vector control ID RHS4080; GFP targeting control clone ID RHS4459; ECSIT clone ID TRCN0000113957 or TRCN0000113958 ; TRAF6 clone ID TRCN0000040733 or TRCN0000040735) and were packaged as described in the Broad Institute RNAi Consortium protocols (http://www.broadinstitute.org/genome_bio/trc/publicProtocols.html) using 239FT cells (Invitrogen). Macrophages were transduced with shRNA lentiviruses overnight on day three of culture and media were changed on day four as above, except that 3 µg ml^{-1} puromycin was added for the remainder of culture to select for transduced cells. On day seven, cells were lifted from plates by incubating in cold TEN buffer (40 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM EDTA), re-plated in fresh media without puromycin containing 10% L929 conditioned media and allowed to rest for a period of 24–48 h before experimentation. RAW cells stably expressing control, *GFP*, *Traf6*, or *Ecsit* shRNAs were selected by puromycin resistance and maintained in 3–4 µg ml^{-1} puromycin.

Ecsit knockdown experiments both in the main Figures and in the Supplementary Figures used shRNA clone ID TRCN0000113958 (termed *Ecsit* sh) except Supplementary Fig. 11, which used clone ID TRCN0000113957 (termed *Ecsit* sh4). All *Traf6* knockdown experiments used shRNA clone ID TRCN0000040735 (termed *Traf6* sh) with the exception of Supplementary Fig. 11, which used clone ID TRCN0000040733 (termed *Traf6* sh1). We observed similar knockdowns and

deficiencies in both mitochondrial and cellular ROS production compared to control when either *Traf6* or *Ecsit* shRNA clone was used but have not included both data sets because of space limitations. Furthermore, we detected no significant differences in ROS production or bacterial killing between macrophages transduced with vector control or GFP control shRNAs; therefore, most of the control samples were transduced with vector control lentiviruses. Finally, we used wild-type BMDMs for control and *Traf6*-knockdown experiments but *Ecsit*$^{+/-}$ BMDMs for *Ecsit* knockdowns. Heterozygous BMDMs have a ~40% reduction in total ECSIT protein abundance by western blot (Supplementary Fig. 11a); therefore, we used these cells to allow for more robust reduction in ECSIT levels upon shRNA silencing. We used age- and sex-matched littermate controls for the wild-type Ctrl and *Traf6* knockdowns, thus minimizing animal to animal variation when comparing wild-type and +/− samples to one another. TRAF6-null BMDMs generation and retroviral reconstitution with WT and RING mutant constructs was performed as previously described[36].

Confocal microscopy. For all microscopy images, cells were grown on coverslips and transfected, stimulated or infected as described. After washing, cells were fixed with 4% paraformaldehyde for 20 min, permeabilized with 0.1% Triton X-100 in PBS for 5 min, blocked with PBS containing 10% FBS for 30 min, stained with primary antibodies for 6 min, then stained with secondary antibodies for 60 min. Cells were washed with PBS between each step. Nuclei were stained with TOPRO3 (Invitrogen) or DAPI (Sigma) and coverslips were mounted with Prolong Gold anti-fade reagent (Molecular Probes). Cells were imaged on a Zeiss LSM 510 META with a 63× water-corrected objective.

Amplex Red assay. 1×10^5 wild-type or *Ecsit*$^{+/-}$ BMDMs infected with control or *Ecsit* shRNA-expressing lentiviruses were plated in 96-well fluorescence microplates (Costar). After two PBS washes to remove residual medium, cells were re-suspended in 100 µl pre-warmed, sterile Hank's buffered salt solution containing 50 µM Amplex Red (Invitrogen) and 1 U ml^{-1} horseradish peroxidase (Sigma) with or without PMA or serum-opsonized *Salmonella* SL1344 for the indicated times. Oxidation of Amplex Red into red-fluorescent resorufin by H_2O_2 was measured in a fluorescence plate reader using excitation at 530 nm and fluorescence detection at 590 nm. Molar concentrations of extracellular peroxide were determined by comparing fluorescence values from each sample with those generated from a standard curve of H_2O_2. Average background fluorescence readings from unstimulated triplicate wells were subtracted from each sample that was stimulated with PMA or infected with *Salmonella*. This allowed for measurement of the amount of H_2O_2 that was induced over unstimulated samples during the same time interval and normalized for cell plating errors.

Nitric Oxide and ELISA assays. Wild-type or *Ecsit*$^{+/-}$ BMDMs were infected with control, *Traf6* or *Ecsit* shRNA-expressing lentiviruses, plated in 24-well dishes and left untreated or stimulated with 1 µg ml^{-1} LPS for the indicated times. Supernatant was collected, serially diluted and measured for nitrite by the Griess Reagent Kit (Molecular Probes) or for TNF-α and IL-12p40 by ELISA (eBioscience) as per the manufacturers' instructions.

31. Kang, B. H. *et al.* Regulation of tumor cell mitochondrial homeostasis by an organelle-specific Hsp90 chaperone network. *Cell* **131,** 257–270 (2007).
32. Hoiseth, S. K. & Stocker, B. A. Aromatic-dependent *Salmonella typhimurium* are non-virulent and effective as live vaccines. *Nature* **291,** 238–239 (1981).
33. Valdivia, R. H. & Falkow, S. Bacterial genetics by flow cytometry: rapid isolation of *Salmonella typhimurium* acid-inducible promoters by differential fluorescence induction. *Mol. Microbiol.* **22,** 367–378 (1996).
34. Cooper, M. P. *et al.* Defects in energy homeostasis in Leigh syndrome French Canadian variant through PGC-1α/LRP130 complex. *Genes Dev.* **20,** 2996–3009 (2006).
35. Sebastiaan Winkler, G. *et al.* Isolation and mass spectrometry of transcription factor complexes. *Methods* **26,** 260–269 (2002).
36. Walsh, M. C., Kim, G. K., Maurizio, P. L., Molnar, E. E. & Choi, Y. TRAF6 autoubiquitination-independent activation of the NFκB and MAPK pathways in response to IL-1 and RANKL. *PLoS ONE* **3,** e4064 (2008).

LETTER

A diverse range of gene products are effectors of the type I interferon antiviral response

John W. Schoggins[1], Sam J. Wilson[2], Maryline Panis[1], Mary Y. Murphy[1], Christopher T. Jones[1], Paul Bieniasz[2] & Charles M. Rice[1]

The type I interferon response protects cells against invading viral pathogens. The cellular factors that mediate this defence are the products of interferon-stimulated genes (ISGs). Although hundreds of ISGs have been identified since their discovery more than 25 years ago[1–3], only a few have been characterized with respect to antiviral activity. For most ISG products, little is known about their antiviral potential, their target specificity and their mechanisms of action. Using an overexpression screening approach, here we show that different viruses are targeted by unique sets of ISGs. We find that each viral species is susceptible to multiple antiviral genes, which together encompass a range of inhibitory activities. To conduct the screen, more than 380 human ISGs were tested for their ability to inhibit the replication of several important human and animal viruses, including hepatitis C virus, yellow fever virus, West Nile virus, chikungunya virus, Venezuelan equine encephalitis virus and human immunodeficiency virus type-1. Broadly acting effectors included IRF1, C6orf150 (also known as MB21D1), HPSE, RIG-I (also known as DDX58), MDA5 (also known as IFIH1) and IFITM3, whereas more targeted antiviral specificity was observed with DDX60, IFI44L, IFI6, IFITM2, MAP3K14, MOV10, NAMPT (also known as PBEF1), OASL, RTP4, TREX1 and UNC84B (also known as SUN2). Combined expression of pairs of ISGs showed additive antiviral effects similar to those of moderate type I interferon doses. Mechanistic studies uncovered a common theme of translational inhibition for numerous effectors. Several ISGs, including ADAR, FAM46C, LY6E and MCOLN2, enhanced the replication of certain viruses, highlighting another layer of complexity in the highly pleiotropic type I interferon system.

To identify novel antiviral effectors in the type I interferon (IFN) system, we developed a cell-based assay to determine the effects of hundreds of ISG products on viral replication. The assay relied on a bicistronic lentiviral vector co-expressing an ISG and the red fluorescent protein TagRFP (Fig. 1a and Supplementary Fig. 1). ISG–TagRFP-expressing cells were challenged with a green fluorescent protein (GFP)-expressing virus, and viral replication within the TagRFP-positive population was quantified by fluorescence-activated cell sorting (FACS) (Fig. 1b). In pilot experiments, 11 ISGs (PSMB8, PSMB9, IFI6, GBP1, IFITM3, MX1, PLSCR1, RTP4, CCDC75, IFIT3 and IFI27) and two control luciferase genes, from Photinus pyralis (Fluc) and Gaussia princeps (Gluc), were tested for the ability to inhibit vesicular stomatitis virus (VSV) and yellow fever virus (YFV) in STAT1-deficient human (STAT1−/−) fibroblasts or in human hepatoma (Huh-7) cells. STAT1−/− fibroblasts produce IFN but do not respond to IFN[4], whereas Huh-7 cells respond to IFN but are not robust producers of IFN[5]. Inhibition was assessed quantitatively by FACS (Fig. 1c) and qualitatively by fluorescence microscopy (Supplementary Fig. 1c). IFI6 and MX1 showed specificity for YFV and VSV, respectively (Supplementary Fig. 1d). Both viruses were inhibited by IFITM3, which also suppresses influenza A virus, West Nile virus (WNV) and dengue virus[6]. These results demonstrate the utility of this

screening platform and suggest its potential for uncovering antiviral ISGs that target divergent virus families.

To conduct a large-scale antiviral ISG screen, we analysed published microarray data sets from IFN-treated cells or tissues, and we established inclusion criteria[1,7–15] (Supplementary Table 1). We compiled a list of 389 ISGs, inserted each into the bicistronic lentiviral vector and produced high-titre lentiviral stocks (Methods). Approximately 50 of these lentiviral stocks failed to mediate high-level transduction of target cells (Supplementary Fig. 2a), perhaps resulting from ISG-mediated toxicity, auto-downregulation of the internal ribosome entry site (IRES)-driven TagRFP expression, poor packaging of the vector genome or activity against the human immunodeficiency virus 1 (HIV-1)-based vector. We consider that the last possibility is likely because several known anti-retroviral genes—EIF2AK2 (also known as PKR), APOBEC3G, BST2 (also known as tetherin) and MOV10—were included in this subset[16–18]. The potential cytotoxicity of ISG overexpression was inferred from FACS-based cell counts (Supplementary Fig. 3a) and validated for a number of genes by luminescence-based cell viability assay (Supplementary Fig. 3b). Although two genes were toxic (CDKN1A and TNFRSF10A), most ISGs did not strongly affect cell viability.

The lentivirus-expressed ISG collection was screened for inhibition of hepatitis C virus (HCV) expressing YPet–GFP (Supplementary Fig. 4) in Huh-7.5 cells, a RIG-I-defective derivative of Huh-7 cells, at 48 h and 72 h after infection. Data are presented as a normal distribution (Fig. 1d) or dot plots (Fig. 1e). Most genes inhibited HCV infection to some extent, with the mean (z score $= 0$) level of inhibition being 20%. Four genes (MDA5, RIG-I, IRF1 and IRF7) had z scores of less than -3.0, whereas 53 genes had z scores between -1.0 and -3.0 (Supplementary Tables 2 and 3). At 72 h after infection, HCV replication approached a maximum (Supplementary Fig. 4d), and the overall profile of ISGs that had antiviral activity was more constricted (Fig. 1e), probably reflecting the ability of the virus to overcome inhibition. Transient transfection was used to screen the subset of ISGs that were poorly expressed following lentiviral transduction, and several genes (including DDX60, MOV10 and MS4A4A) were found to have anti-HCV activity (Supplementary Fig. 5). To investigate the effect of the host cell environment on ISG activity, we rescreened HCV in a less-permissive cell type (Huh-7)[19]. The strongest effectors were consistently active in both cell lines (Fig. 1e and Fig. 2a), and several additional genes (including MAP3K14 and SLC1A1) inhibited HCV replication to a greater extent in Huh-7 cells than in Huh-7.5 cells (z score less than -2.0; Supplementary Table 4). Taken together, these data suggest that there are at least two categories of anti-HCV ISGs: strong inhibitors that probably feed back into IFN-mediated signalling pathways or other signalling pathways, and modest inhibitors that may have targeted effector functions.

We also screened the ISG collection against other medically important viruses from three families: Flaviviridae, YFV and WNV; Togaviridae, chikungunya virus (CHIKV) and Venezuelan equine encephalitis virus (VEEV); and Retroviridae, HIV-1 (Fig. 2a). We

[1]Laboratory of Virology and Infectious Disease, Center for the Study of Hepatitis C, The Rockefeller University, New York, New York 10065, USA. [2]Howard Hughes Medical Institute, Laboratory of Retrovirology, Aaron Diamond AIDS Research Center, The Rockefeller University, New York, New York 10016, USA.
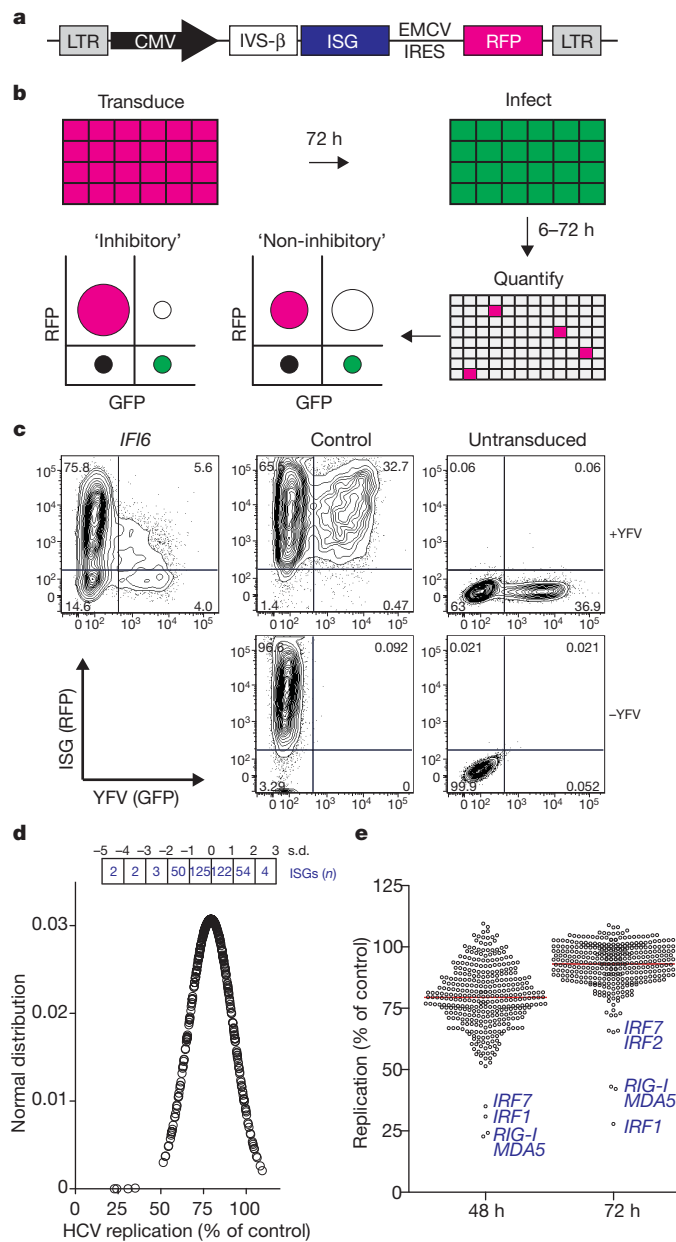
Figure 1 | FACS-based screen for identifying antiviral ISGs. a, Schematic of the Gateway-compatible, bicistronic lentiviral vector. CMV, immediate early promoter from human cytomegalovirus; EMCV, encephalomyocarditis virus; IVS-β, rabbit β-globin gene intron II; LTR, HIV-1 long terminal repeat. **b**, Schematic of the overexpression screen showing hypothetical FACS plots for cells infected with lentiviral vectors expressing inhibitory (antiviral) genes and non-inhibitory genes. The overlap of RFP (red) and GFP (green) is depicted as white. **c**, FACS plots showing *IFI6*-mediated inhibition of YFV in *STAT1*$^{-/-}$ fibroblasts. The units for the x and y axes are fluorescence intensity of GFP and RFP, respectively. The numbers in each quadrant indicate the percentage of the total cell population. **d**, Distribution of HCV replication levels in the presence of overexpressed ISGs at 48 h, normalized to the Fluc control. The number (n) of ISGs within standard deviation (s.d. or z score) ranges is shown in the boxes. **e**, Dot plots of HCV replication levels in the presence of overexpressed ISGs at 48 h and 72 h, normalized to Fluc control. Selected ISGs are indicated in blue. The red line indicates the population mean.

assayed YFV, WNV, VEEV and CHIKV in *STAT1*$^{-/-}$ fibroblasts and HIV-1 in MT-4 cells (a human T lymphotrophic virus I (HTLV-I)-transformed T-cell line). Global ISG inhibition profiles were similar for all viruses, with 10–25 genes having z scores of −1.5 or lower (Fig. 2a and Supplementary Tables 7–10). YFV was also screened in Huh-7 cells (Supplementary Fig. 6a), and this screen yielded fewer hits,

with five genes showing a z score below −1.5 (Supplementary Tables 5 and 6). The strongest hits in Huh-7 cells (*HPSE*, *IRF1*, *IRF7*, *IFITM3* and *IFI6*) were also among the top inhibitors in *STAT1*$^{-/-}$ fibroblasts, suggesting that these effectors function independently of cell type. Differences in data between the two cell lines may be due to higher levels of YFV replication in Huh-7 cells (Supplementary Fig. 6b) and/or increased ISG expression in *STAT1*$^{-/-}$ fibroblasts (Supplementary Fig. 2b). These differences indicate that *STAT1*$^{-/-}$ fibroblasts provide a more sensitive environment for detecting ISG-mediated inhibition.

The strongest antiviral ISGs from each primary screen were validated by generating independent lentiviral stocks and measuring inhibition from at least eight replicates (Fig. 2b). Forty-five hits (ISGs) were confirmed to reduce viral replication significantly compared with a control gene (Fluc) ($P < 0.001$ for 44 genes and $P < 0.05$ for 1 gene), and five genes were false positives. Further validation was conducted for a subset of ISGs. Inhibition of YFV by *IFI6*, *IFITM3* and *RTP4* was dose dependent (Supplementary Fig. 7a), and both *IFITM3* and *IFI6* inhibited production of the parental, non-GFP-expressing, virus (YFV-17D) (Supplementary Fig. 7b, c). We also measured the effects of anti-HCV ISGs when they were introduced after infection with HCV (Supplementary Fig. 8). In contrast to the 20–70% range in inhibition observed when ISGs were expressed before infection (Fig. 2b), ISG expression after infection affected replication by either less than 20% (for *DDIT4*, *NT5C3*, *IFI44L*, *MAP3K14* and *IRF2*) or more than 80% (for *IRF7*, *IRF1*, *MDA5* and *RIG-I*). Only one gene, *OASL*, showed an intermediate phenotype (Supplementary Fig. 8). These data demonstrate that the overexpression platform is highly reliable for identifying antiviral ISGs and is sensitive enough to detect even modest levels of inhibition both before and after infection.

In the primary screens, several genes enhanced viral replication (Fig. 2a). This activity was confirmed for six of these (*ADAR*, *APOBEC3A*, *FAM46C*, *IDO1*, *LY6E* and *MCOLN2*) by testing against YFV, WNV, VEEV and CHIKV in *STAT1*$^{-/-}$ fibroblasts (Fig. 2d). For most of these six genes, we found an increase in the number of YFV- or VEEV-infected cells, whereas the replication of WNV and CHIKV was largely unaffected. Consistent with the primary YFV screen, *LY6E* conferred a striking 75% increase in infection frequency (Fig. 2d and Supplementary Fig. 9), a phenotype that may correlate with studies implicating *LY6E* in the susceptibility of chickens to Marek's disease virus[20] and of mice to adenovirus[21]. Analysis of mean GFP intensity in infected cells revealed that *ADAR* significantly enhanced the replication of all viruses against which it was tested (Supplementary Fig. 9). Thus, in contrast to *LY6E*, which caused an increase in the number of infected cells, *ADAR* affected the level of replication within the cell. *ADAR* encodes an RNA-specific adenosine deaminase that has been reported to stimulate HIV-1 infection through RNA editing[22]. *ADAR* also promotes infection with measles virus[23], supporting a pan-viral mechanism of enhancement.

The screening and validation data uncovered both broad-acting effectors and specific effectors. *IRF1*, which encodes a transcription factor in the IFN pathway, inhibited all viruses against which it was tested (Fig. 2c), even in a *STAT1*$^{-/-}$ background. Given that there are more than 200 IRF1-binding sites in the human genome[24], our data suggest that IRF1 activates a unique antiviral program. Indeed, microarray analysis of IRF1-transduced Huh-7 cells or *STAT1*$^{-/-}$ fibroblasts showed that numerous ISGs were induced (Supplementary Fig. 10), but no IFN genes were upregulated (data not shown). Most genes that were induced greater than threefold in Huh-7 cells ($P < 0.05$) were similarly upregulated in *STAT1*$^{-/-}$ fibroblasts, and both of these cell lines had a subset of uniquely regulated transcripts. Several of our antiviral hits were among the IRF1-induced transcripts, supporting the observation that the effector mechanisms of IRF1 partly overlap with those of IFN but do not require IFN activity. The genes *C6orf150*, *HPSE*, *RIG-I*, *MDA5*, *IFITM3*, *IRF7* and *NAMPT* were also active against more than one viral species, whereas other genes showed more restricted antiviral specificities. Gene ontology analysis
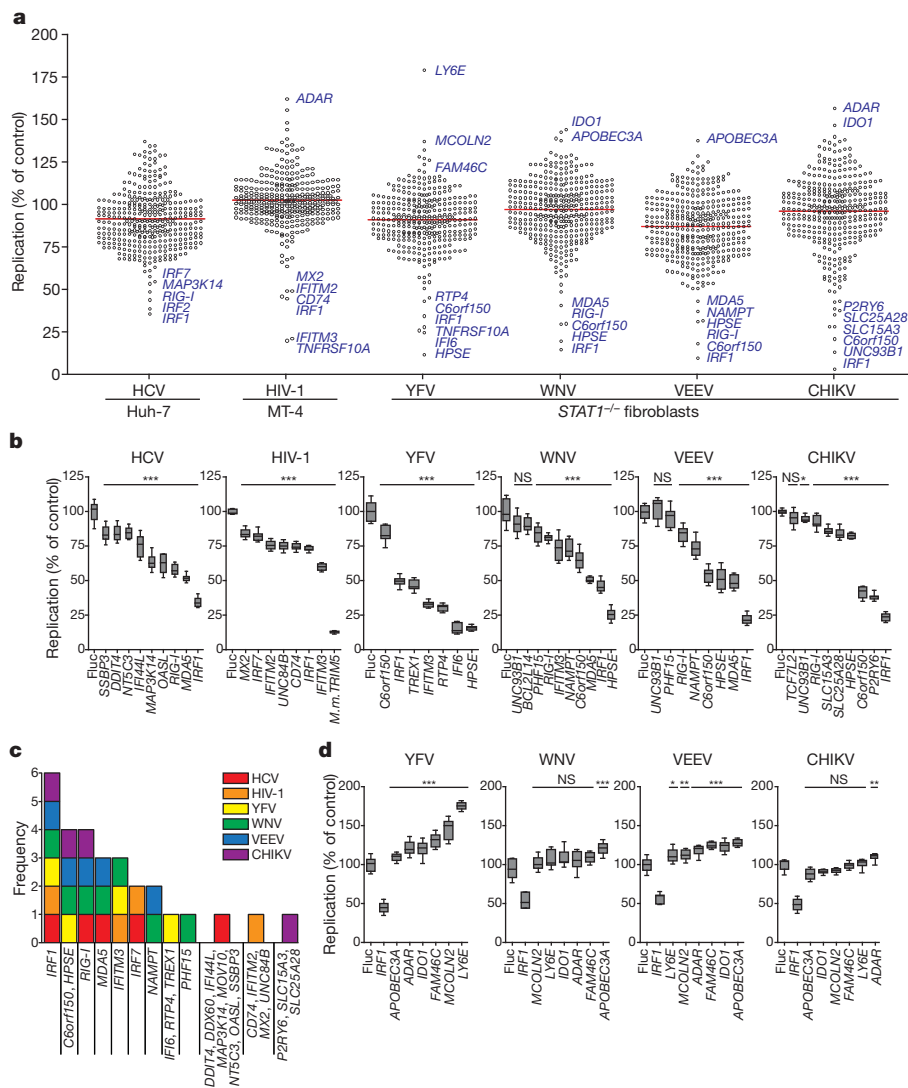
Figure 2 | Identification of ISGs that inhibit or enhance viral replication.
a, Dot plots of large-scale ISG screens against six viruses. Replication levels were normalized to the Fluc control. Selected ISGs are indicated in blue. The red line indicates the population mean. **b, d**, Confirmation assays for selected inhibitory (**b**) and enhancing (**d**) ISGs. For HIV-1, *Macaca mulatta* (*M.m.*) *TRIM5* was included as a control. Replication levels were normalized to the Fluc control.

Data are presented as box and whisker plots: grey boxes extend from the 25th to the 75th percentile, with a black line at the population median; whiskers extend to show the highest and lowest values ($n = 8$ for HCV and $n = 9$ for other viruses). Statistical significance was determined by one-way ANOVA (***, $P < 0.001$; **, $P < 0.01$; *, $P < 0.05$; NS, not significant). **c**, Frequency of validated antiviral ISG activity across six screens.

was performed to classify the confirmed antiviral ISGs. Compared with the entire ISG collection, the validated list had a significant over-representation of three molecular functions (nucleic acid binding, hydrolase activity and helicase activity) and three biological processes (signal transduction, transcription initiation and small molecule transport) (Supplementary Fig. 11 and Supplementary Table 11).

The identification of modest to strong virus inhibitors suggests that an effective IFN response requires the combinatorial action of numerous ISGs. This hypothesis was initially proposed from microarray[1] and gene knockout studies[25], but there are limited experimental data showing combinatorial ISG function. To determine whether the effects of ISGs are greater when the genes are expressed in combination, selected hits from the HCV, HIV-1 and YFV screens were tested in all of the possible two-gene pairs (Fig. 3a and Supplementary Fig. 12). ISG effects were generally additive, with some combinations reducing replication by greater than 90%, an efficacy similar to moderate doses of IFN (Fig. 3b–e). For YFV, when one inhibitory and one enhancing ISG were tested together (for example, *C6orf150* and *MCOLN2*), the magnitude of viral replication was more strongly influenced by the inhibitory gene.

To probe the mechanisms of antiviral action, selected ISGs were tested for the ability to inhibit various stages of the viral life cycle.

Using HCV pseudoparticles, none of the eight anti-HCV ISGs significantly impaired virus entry into Huh-7 cells (Supplementary Fig. 13a). By contrast, analysis of subgenomic HCV replicons expressing the Gluc reporter showed that all ISGs that were tested inhibited primary translation by 25–70% at 4 h post transfection (Fig. 4a, b and Supplementary Fig. 13b, c). Notably, the strongest translational inhibitors were also the most potent suppressors of replication, as measured by Gluc protein production at 72 h (Fig. 4b) and confirmed by quantifying the HCV genome copy number (Supplementary Fig. 13d). The finding that each of the eight genes, which have a range of predicted molecular functions (Supplementary Table 11), blocked initial translation suggests the importance of this host cell strategy for suppressing HCV. *IRF1* and *RIG-I* were also associated with a kinetic delay in the onset of replication, suggesting that there are multiple possible stages at which the HCV life cycle can be blocked by these effectors.

We also investigated the mechanism of action of *C6orf150*, which encodes a putative nucleotidyltransferase[26] and contains a Mab-21 domain but is otherwise uncharacterized. Because *C6orf150* inhibited the alphaviruses CHIKV and VEEV, we tested it against the related virus Sindbis virus (SINV). *C6orf150* inhibited the replication of GFP-expressing SINV (Fig. 4c) and infectious virus production by
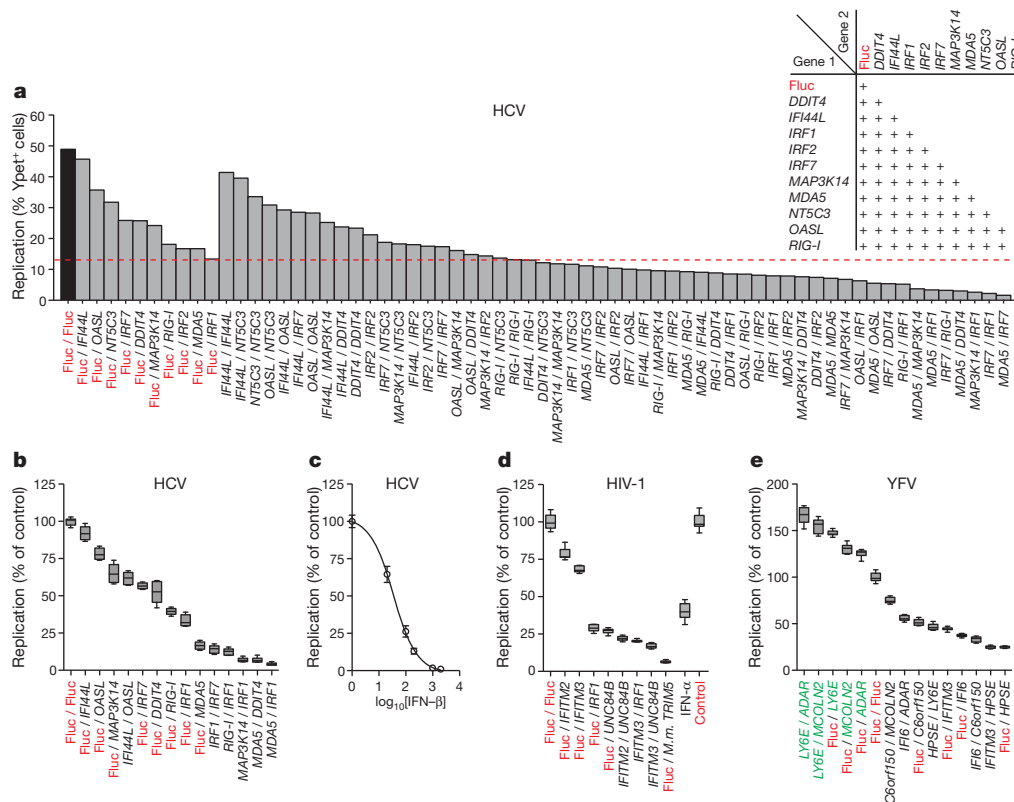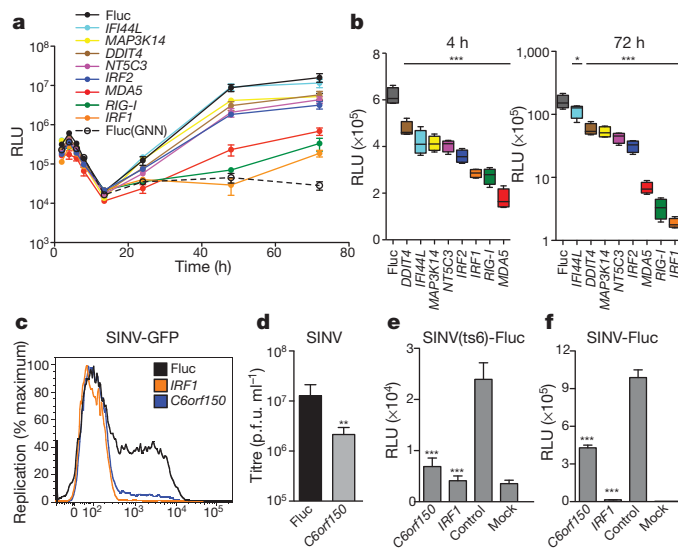
Figure 3 | Combinatorial action of inhibitory and enhancing ISGs. a, Anti-HCV ISGs were tested in two-gene combinations (inset). The dashed red line denotes the level of viral replication achieved in the presence of the strongest single-gene inhibitor, *IRF1*. **b, d, e**, Confirmation assays from two-gene screens, against HCV (**b**), HIV-1 (**d**) and YFV (**e**). ISGs are colour coded: Fluc control

(red), inhibitory (black) and enhancing (green). Data were normalized to the Fluc control ($n = 8$ for HCV and YFV, and $n = 9$ for HIV-1). Replication of HIV-1 in IFN-α-treated (black) or untreated (red) cells is plotted for comparison. **c**, Inhibition of HCV by IFN-β. Results are presented as mean ± s.d. ($n = 3$).

non-GFP-expressing SINV (Fig. 4d). We next assessed the activity of *C6orf150* against a Fluc-expressing temperature-sensitive virus, SINV(ts6), which is competent for primary translation but does not replicate at non-permissive temperatures (Methods). SINV(ts6) was susceptible to inhibition by *C6orf150* (Fig. 4e), indicating an early viral translation block. Interestingly, the pan-viral inhibitor IRF1 also blocked SINV(ts6) translation, suggesting that this effector mechanism is reinforced by IFN-induced transcriptional factors. Similar to the HCV life-cycle studies, the importance of translational inhibition is highlighted by a corresponding reduction in replication of wild-type SINV-Fluc (Fig. 4f).

The antiviral effects of type I IFN were first described more than 50 years ago[27]. Since then, great strides have been made in understanding the mechanisms that govern virus-mediated IFN production, IFN gene regulation and IFN-mediated signalling[28,29]. Insight into downstream IFN effector mechanisms, however, has largely been limited to a select group of ISGs, probably because of the difficulties involved in systematically overexpressing hundreds of genes. With the strategy presented here, we have overcome these technical barriers and identified multiple novel antiviral ISGs, many of which target early steps in the viral life cycle. Combined with our previous understanding of well-known antiviral effectors such as PKR (which is encoded by *EIF2AK2*) and IFIT proteins, life-cycle stage characterization of these newly identified ISGs suggests that the IFN system may use multiple strategies to engender a single outcome such as translational inhibition. Clinically, IFN is used to treat several viral infections and is a principle component of HCV treatment. Given the clinical side effects of IFN treatment, and the natural ability of viruses to thwart its activity[30], exploiting the actions

Figure 4 | Translational inhibition is a common mechanism of ISG-mediated antiviral action. a, ISG-expressing Huh-7 cells were transfected with HCV subgenomic replicon RNA encoding Gluc protein, and the Gluc relative light units (RLU) in the cell supernatants were measured. Results are presented as mean ± s.d. ($n = 4$). GNN, a replication-defective HCV replicon. **b**, Inhibition of primary translation (left) or replication (right) was inferred from the 4 h and 72 h data from **a**, respectively. Results are presented as mean ± s.d. ($n = 4$). Statistical significance was determined by one-way ANOVA (***, $P < 0.001$; *, $P < 0.05$). **c–f**, *C6orf150*-mediated inhibition of SINV in *STAT1*[−/−] fibroblasts. *C6orf150* inhibits the replication of SINV-GFP (**c**) and the production of non-GFP-expressing SINV (**d**). *C6orf150* and *IRF1* inhibit the primary translation of SINV(ts6)-Fluc (**e**) and the replication of SINV-Fluc (**f**). Results are presented as mean ± s.d. ($n = 3$). **c–f**, Statistical significance was determined by one-way ANOVA (***, $P < 0.001$; **, $P < 0.01$). p.f.u., plaque-forming units.

of individual ISGs may be a preferable therapeutic strategy. Further insight into the biochemical mechanisms of these effectors may provide a platform for the development of alternatives to IFN-based therapies.

METHODS SUMMARY

To screen ISGs for antiviral activity, we obtained sequence-validated ORF*EXPRESS* ISG shuttle clones (GeneCopoeia) and used Gateway technology to move genes into a lentiviral vector co-expressing TagRFP (Evrogen). Lentiviruses were packaged in 293T cells by using established techniques (Methods). Target cell lines were transduced with the lentiviral ISG stocks for 2–3 days and then infected with a GFP-expressing virus at a multiplicity of infection of approximately 0.5–1.0. Cells were collected in Accumax Cell Aggregate Dissociation Medium (eBioscience), fixed in 1% paraformaldehyde and analysed on an LSRII-HTS flow cytometer (BD Biosciences). To quantify viral replication, RFP^+ cells were gated, and the number of GFP^+ cells in the RFP^+ population was determined. Replication levels were normalized to controls (Fluc or Gluc) to obtain z scores. In some cases, we measured replication levels by determining mean GFP fluorescence intensity in the RFP^+GFP^+ population. For all ISG hit validation experiments, we generated lentiviral stocks that were independent from the primary screening preparations. To determine statistical significance, we used one-way analysis of variance (ANOVA) for at least eight replicates. For HCV life-cycle experiments, ISG-expressing Huh-7 cells were infected with genotype 1b HCV pseudoparticles or transfected with *in vitro*-transcribed HCV subgenomic RNA expressing Gluc. For SINV life-cycle experiments, ISG-expressing $STAT1^{-/-}$ fibroblasts were infected with SINV variants, and cells were monitored for GFP levels by FACS, for Fluc levels by reporter assay and for virus production by plaque assay on BHK-J cells. Cell viability and reporter gene assays for Gluc and Fluc were performed with commercially available kits (Promega). Gene ontology analysis was performed using the PANTHER Classification System (http://www.pantherdb.org). Microarray studies were performed on total RNA collected from Fluc-expressing or IRF1-expressing cells using BeadArray technology (Illumina).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. de Veer, M. J. *et al.* Functional classification of interferon-stimulated genes identified using microarrays. *J. Leukoc. Biol.* **69,** 912–920 (2001).
2. Knight, E. Jr & Korant, B. D. Fibroblast interferon induces synthesis of four proteins in human fibroblast cells. *Proc. Natl Acad. Sci. USA* **76,** 1824–1827 (1979).
3. Larner, A. C. *et al.* Transcriptional induction of two genes in human cells by β interferon. *Proc. Natl Acad. Sci. USA* **81,** 6733–6737 (1984).
4. Dupuis, S. *et al.* Impaired response to interferon-α/β and lethal viral disease in human STAT1 deficiency. *Nature Genet.* **33,** 388–391 (2003).
5. Keskinen, P. *et al.* Impaired antiviral response in human hepatoma cells. *Virology* **263,** 364–375 (1999).
6. Brass, A. L. *et al.* The IFITM proteins mediate cellular resistance to influenza A H1N1 virus, West Nile virus, and dengue virus. *Cell* **139,** 1243–1254 (2009).
7. Brodsky, L. I. *et al.* A novel unsupervised method to identify genes important in the anti-viral response: application to interferon/ribavirin in hepatitis C patients. *PLoS ONE* **2,** e584 (2007).
8. He, X. S. *et al.* Global transcriptional response to interferon is a determinant of HCV treatment outcome and is modified by race. *Hepatology* **44,** 352–359 (2006).
9. Hilkens, C. M., Schlaak, J. F. & Kerr, I. M. Differential responses to IFN-α subtypes in human T cells and dendritic cells. *J. Immunol.* **171,** 5255–5263 (2003).
10. Hultcrantz, M. *et al.* Interferons induce an antiviral state in human pancreatic islet cells. *Virology* **367,** 92–101 (2007).
11. Indraccolo, S. *et al.* Identification of genes selectively regulated by IFNs in endothelial cells. *J. Immunol.* **178,** 1122–1135 (2007).
12. Lanford, R. E. *et al.* Genomic response to interferon-α in chimpanzees: implications of rapid downregulation for hepatitis C kinetics. *Hepatology* **43,** 961–972 (2006).
13. Leaman, D. W. *et al.* Novel growth and death related interferon-stimulated genes (ISGs) in melanoma: greater potency of IFN-β compared with IFN-α2. *J. Interferon Cytokine Res.* **23,** 745–756 (2003).
14. Rani, M. R. *et al.* Novel interferon-β-induced gene expression in peripheral blood cells. *J. Leukoc. Biol.* **82,** 1353–1360 (2007).
15. Sarasin-Filipowicz, M. *et al.* Interferon signaling and treatment outcome in chronic hepatitis C. *Proc. Natl Acad. Sci. USA* **105,** 7034–7039 (2008).
16. Roy, S. *et al.* Control of the interferon-induced 68-kilodalton protein kinase by the HIV-1 *tat* gene product. *Science* **247,** 1216–1219 (1990).
17. Neil, S. & Bieniasz, P. Human immunodeficiency virus, restriction factors, and interferon. *J. Interferon Cytokine Res.* **29,** 569–580 (2009).
18. Furtak, V. *et al.* Perturbation of the P-body component Mov10 inhibits HIV-1 infectivity. *PLoS ONE* **5,** e9081 (2010).
19. Blight, K. J., McKeating, J. A. & Rice, C. M. Highly permissive cell lines for subgenomic and genomic hepatitis C virus RNA replication. *J. Virol.* **76,** 13001–13014 (2002).
20. Liu, H. C., Niikura, M., Fulton, J. E. & Cheng, H. H. Identification of chicken lymphocyte antigen 6 complex, locus E (*LY6E*, alias *SCA2*) as a putative Marek's disease resistance gene via a virus–host protein interaction screen. *Cytogenet. Genome Res.* **102,** 304–308 (2003).
21. Spindler, K. R. *et al.* The major locus for mouse adenovirus susceptibility maps to genes of the hematopoietic cell surface-expressed LY6 family. *J. Immunol.* **184,** 3055–3062 (2010).
22. Doria, M., Neri, F., Gallo, A., Farace, M. G. & Michienzi, A. Editing of HIV-1 RNA by the double-stranded RNA deaminase ADAR1 stimulates viral infection. *Nucleic Acids Res.* **37,** 5848–5858 (2009).
23. Toth, A. M., Li, Z., Cattaneo, R. & Samuel, C. E. RNA-specific adenosine deaminase ADAR1 suppresses measles virus-induced apoptosis and activation of protein kinase PKR. *J. Biol. Chem.* **284,** 29350–29356 (2009).
24. Frontini, M., Vijayakumar, M., Garvin, A. & Clarke, N. A. ChIP-chip approach reveals a novel role for transcription factor IRF1 in the DNA damage response. *Nucleic Acids Res.* **37,** 1073–1085 (2009).
25. Zhou, A., Paranjape, J., Der, S. D., Williams, B. R. & Silverman, R. H. Interferon action in triply deficient mice reveals the existence of alternative antiviral pathways. *Virology* **258,** 435–440 (1999).
26. Kuchta, K., Knizewski, L., Wyrwicz, L. S., Rychlewski, L. & Ginalski, K. Comprehensive classification of nucleotidyltransferase fold proteins: identification of novel families and their representatives in human. *Nucleic Acids Res.* **37,** 7701–7714 (2009).
27. Isaacs, A. & Lindenmann, J. Virus interference. I. The interferon. *Proc. R. Soc. Lond. B* **147,** 258–267 (1957).
28. Wilkins, C. & Gale, M. Jr. Recognition of viruses by cytoplasmic sensors. *Curr. Opin. Immunol.* **22,** 41–47 (2010).
29. Takaoka, A. & Yanai, H. Interferon signalling network in innate defence. *Cell. Microbiol.* **8,** 907–922 (2006).
30. Bowie, A. G. & Unterholzner, L. Viral evasion and subversion of pattern-recognition receptor signalling. *Nature Rev. Immunol.* **8,** 911–922 (2008).

METHODS

Viruses, replicons and cells. Huh-7, Huh-7.5 and 293T cells were maintained in DMEM (Invitrogen) with 10% FBS and 0.1 mM non-essential amino acids. $STAT1^{-/-}$ fibroblasts (an SV40 large T antigen immortalized skin fibroblast line) and MT-4 cells (an HTLV-I immortalized T-cell line) were grown in RPMI (Invitrogen) with 10% FBS. BHK-J cells were grown in MEM (Invitrogen) with 7.5% FBS. The recombinant infectious viruses used were as follows: HCV-YPet[31] (derived from Bi-YPet-Jc1FLAG2), YFV-Venus[31] (derived from YF17D-5′C25Venus2AUbi), WNV-GFP[32] (derived from pBELO-WNV-GFP-RZ ic), CHIKV-GFP[33] (derived from pCHIKV-LR 5′GFP), SINV/SINV-GFP[34] (derived from pS300/pS300–GFP), and SINV-Fluc and the temperature-sensitive variant SINV(ts6)-Fluc (derived from pToto1101/Luc and pToto11-1/Luc:ts6). SINV(ts6)-Fluc carries mutations that prevent replication at non-permissive temperatures, but the virus retains the ability to translate its incoming genome[35]. VEEV-GFP is a double subgenomic enhanced GFP (EGFP) reporter virus derived from the TC83 vaccine strain of VEEV (provided by I. Frolov).

HCV subgenomic replicon RNAs were derived from Bi-Gluc-JFH(SG) and Bi-Gluc-JFH-GNN(SG), which are bicistronic constructs expressing Gluc from the HCV IRES and genotype 2a JFH-1 non-structural proteins from the EMCV IRES. The GNN variant encodes polymerase mutations and is not competent for replication. Viral stocks were generated by electroporation of in vitro-transcribed RNA into Huh-7.5 cells (for HCV) or BHK-J cells (for YFV, WNV, VEEV, CHIKV and SINV) as previously described[32,33,35–38]. Single-cycle HIV-1-GFP was generated by transfection of HEK 293T cells with a modified env-proviral plasmid encoding GFP in place of nef (pR7/Δenv/GFP)[39]. HIV-1 particles were pseudotyped with the VSV glycoprotein (VSV-G). Lentiviral pseudoparticles with no envelope or carrying HCV glycoprotein were generated and assayed for viral entry as previously described[40]. Experiments with WNV, CHIK, and HIV-1 were carried out under biosafety level 3 containment, in compliance with institutional and federal guidelines.

DNA constructs. To generate a lentivirus-based, Gateway-compatible destination vector for ISG expression, a Gateway expression cassette (containing the tetracycline-inducible-hybrid CMV promoter, chloramphenicol resistance gene and ccdB suicide gene) from pLenti4.TO.V5-DEST (Invitrogen) was subcloned into the XhoI–NdeI sites of pTRIP–EGFP[41]. Overlap extension PCR followed by a three-fragment ligation was used to insert the EMCV IRES and TagRFP-encoding sequence (Evrogen) downstream of the Gateway module (EMCV IRES 5′ oligonucleotide, 5′-GATATCTCGAGGCCCCTCTCCCTCCCCCCCCCCTAA-3′; EMCV IRES 3′ oligonucleotide, 5′-CACGATGATAATATGGCCACAACCCCGCGGATATG-3′; TagRFP 5′ oligonucleotide, 5′-CATAGCTAGCATGGTGTCTAAGGGCGAAGAGCTG; TagRFP 3′ oligonucleotide, 5′-CTAGCAAACTGGGGCACAAACTTAATTGACCGCGGGGTACCTGCG-3′).

To enhance gene expression, the β-globin gene intron IVS-β was subcloned from pLenti6-TR (Invitrogen) into the SpeI site between the CMV promoter and the Gateway cassette (IVS-β 5′ oligonucleotide, 5′-GACCCACTAGTGTGAGTTTGGGGACCCTTGATTG-3′; IVS-β 3′ oligonucleotide, 5′-CATGCCTTCTTCTTTTTCCTACAGACTAGTCCCAG-3′).

All DEST vector variants were grown in DB3.1 cells (Invitrogen) under ampicillin and chloramphenicol double selection. The final vector was named pTRIP.CMV.IVSb.ires.TagRFP-DEST. Sequence-validated, Gateway-compatible ORFEXPRESS shuttle clones corresponding to 387 ISGs were obtained from GeneCopeia. Two additional ISG entry clones, encoding RSAD2 and ZC3HAV1, and two control clones, encoding Fluc and Gluc, were generated as follows. Genes encoding RSAD2, ZC3HAV1, Fluc and Gluc were PCR amplified with oligonucleotides containing attB sites flanking gene-specific sequences (RSAD2 5′ oligonucleotide, 5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACCATGTGGGTGCTTACACCTGC-3′; RSAD2 3′ oligonucleotide, 5′-GGGGACCACTTTGTACAAGAAAGCTGGGTCTACCAATCCAGCTTCAGAT-3′; ZC3HAV1 5′ oligonucleotide, 5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACCATGGCGGACCCGGAGGTGTG-5′; ZC3HAV1 3′ oligonucleotide, 5′-GGGGACCACTTTGTACAAGAAAGCTGGGTTTACTCTGGCCCTCTCTTCATCT-3′; Fluc 5′ oligonucleotide, 5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACCATGGAAGATGCCAAAAACATTAAGAA-3′; Fluc 3′ oligonucleotide, 5′-GGGGACCACTTTGTACAAGAAAGCTGGGTTTACACGGCGATCTTGCCGCCCTTC-3′; Gluc 5′ oligonucleotide, 5′-GGGGACAAGTTTGTACAAAAAAGCAGGCTTCACCATGGGAGTCAAAGTTCTGTTTGCCC-3′; Gluc 3′ oligonucleotide, 5′-GGGGACCACTTTGTACAAGAAAGCTGGGTTTAGTCACCACCGGCCCCCTTGATC-3′).

PCR products were purified over GFX columns (GE Healthcare) and cloned into pDONR (Invitrogen) with BP Clonase. BP Clonase reactions were transformed into OmniMAX-competent Escherichia coli (Invitrogen), and colonies were screened by restriction digestion and sequencing.

The ISG-encoding sequences from pENTR clones were moved into pTRIP.CMV.IVSb.ires.TagRFP-DEST using LR Clonase II (Invitrogen) according to the manufacturer's instructions. LR reaction products were transformed into MDS42Rec reduced genome E. coli (Scarab Genomics). One or two colonies for each clone were grown in 3 ml Luria-Bertani (LB) broth with ampicillin, and transfection-quality plasmid DNA was purified over anion-exchange columns (Qiagen). All pTRIP.CMV.IVSb.ISG.ires.TagRFP constructs were sequenced at the 5′ end of the expression cassette to verify gene insertion (sequencing oligonucleotide, 5′-CCTGCCTTTCTCTTTATGG-3′).

Generation of ISG-expressing lentiviral pseudoparticles. Lentiviral pseudoparticles were generated by co-transfecting 4×10^5 293T cells in 6-well plates or 1.2×10^4 cells in 96-well plates with plasmids expressing the pTRIP.CMV.IVSb.ISG.ires.TagRFP proviral DNA, HIV-1 gag–pol and VSV-G in a ratio of 1/0.8/0.2, respectively. For each transfection, 6 μl FuGENE (Roche) was combined with 2.0 μg total DNA in 100 μl Opti-MEM (Gibco). Transfections were carried out for 6 h, followed by a medium change to DMEM containing 3% FBS. Supernatants were collected at 48 h and 72 h, pooled, cleared by centrifugation and stored at −80 °C.

Lentiviral transduction and viral replication assays. Transduction assays. Huh-7 cells, Huh-7.5 cells and $STAT1^{-/-}$ fibroblasts were seeded into 24-well plates at a density of 7×10^4 cells well^{-1} and transduced with lentiviral pseudoparticles by spinoculation at 1,000–1,500g for 1 h at 37 °C in medium containing 3% FBS, 20 mM HEPES and 4 μg ml^{-1} polybrene. For HIV-1 studies, MT-4 cells in suspension were transduced in a 96-well format. For monolayer experiments with Huh-7 cells, Huh-7.5 cells and $STAT1^{-/-}$ fibroblasts, cells were split 1:2 or 1:3 at 48 h post transduction.

Replication assays. Transduced cells were infected with the indicated virus at a dose yielding approximately 50% infected (GFP$^+$) cells, as previously determined by FACS-based infectivity assays. Infected cells were collected at various time points, typically within the first viral replication cycle when possible: HCV (48 h or 72 h), HIV-1 (48 h), YFV (24 h), WNV (6 h), VEEV (5.5 h) and CHIKV (10 h). Adherent cells were collected into 200 μl Accumax Cell Aggregate Dissociation Medium (eBioscience) and transferred to a 96-well plate. Cells were pelleted at 1,000g for 5 min at 4 °C and resuspended in 1% paraformaldehyde fixation solution, in which they remained for at least 1 h. Cells were then pelleted by centrifugation at 1,000g for 5 min at 4 °C, resuspended in cold 1× PBS containing 3% FBS and stored at 4 °C until FACS analysis was carried out. Samples were analysed in a 96-well-based high-throughput manner using an LSRII-HTS flow cytometer (BD Biosciences) equipped with a 561-nm laser for detection of TagRFP. Data were analysed using FlowJo software (Treestar) with a 0.1% compensation matrix.

For IFN titrations, Huh-7 cells or MT-4 T cells were treated with varying doses of IFN-β (PBL InterferonSource) or 1000 U ml^{-1} IFN-α (Sigma), respectively, for 24 h before infection. Cells were infected with HCV (Huh-7 cells) or HIV-1 (MT-4 cells), and replication levels were monitored by FACS as described above.

For SINV-Fluc and SINV(ts6)-Fluc, cells were infected at 28 °C (ts6 mutant) or 37 °C (wild-type SINV) and collected 4 h post infection. Intracellular Fluc levels were assessed by the Luciferase Assay System reporter assay (Promega) according to the manufacturer's instructions.

For HCV life-cycle studies, subgenomic RNAs were generated by in vitro transcription using the T7 RiboMAX transcription kit (Promega). RNA (175 ng) was transfected into 3.5×10^4 ISG-expressing Huh-7 cells using Mirus reagent (Mirus Bio). Translation and replication were monitored by sampling cell supernatants and assaying Gluc production over time (at 2, 4, 6, 8, 13.5, 24, 48 and 72 h post transfection) with the Renilla Luciferase Assay System (Promega) or by quantifying HCV genome copy number by quantitative reverse transcription–PCR as previously described[42].

DNA-transfection-based ISG screen of Huh-7 cells. Huh-7 cells were seeded into 24-well plates at a density of 7×10^4 cells well^{-1}. The next day, growth medium was changed to DMEM containing 1% FBS and 0.1 mM non-essential amino acids. Lentiviral plasmid DNAs were transfected at 400 ng DNA well^{-1} using Lipofectamine 2000 (Invitrogen) in a total volume of 1 ml well^{-1}. Plates were centrifuged at 1,000g for 30 min at 37 °C. Five hours later, the medium was changed to DMEM containing 10% FBS. At two days post transfection, cells were challenged with HCV-YPet and assayed for replication by using FACS as described above.

Immunofluorescence. $STAT1^{-/-}$ fibroblasts expressing IRF1.ires.TagRFP, C6orf150.ires.TagRFP or MAP3K14.ires.TagRFP were stained to examine protein levels by using Cellomics Whole Cell Stain kit (Thermo Scientific) according to the manufacturer's instructions. Affinity-purified polyclonal rabbit antisera detecting human IRF1, MAP3K14 and C6orf150 were obtained from Abcam or Sigma and were used at the manufacturer's recommended concentrations. An Alexa-Fluor-488-conjugated goat anti-rabbit antibody (Invitrogen) was used at a concentration of 1/1000 for detection, and samples were visualized by epifluorescence microscopy using an Eclipse TE300 inverted microscope (Nikon).

Microarray analysis. Huh-7 cells and $STAT1^{-/-}$ fibroblasts were transduced with lentiviruses expressing Fluc.ires.TagRFP or IRF1.ires.TagRFP. Total RNA was

collected from transduced cells 48 h later using an RNeasy Mini Kit (Qiagen). RNA amplification was carried out with a MessageAmp Premier RNA Amplification Kit (Applied Biosystems). Total RNA (200 ng) was used to synthesize the first strand of cDNA using ArrayScript reverse transcriptase and an oligo(dT) primer bearing a T7 promoter. The single-stranded cDNA was then converted into double-stranded DNA by DNA polymerase I in the presence of *E. coli* RNase H and DNA ligase. The double-stranded DNA served as a template for *in vitro* transcription in a reaction containing biotin-labelled UTP, unlabelled NTPs and T7 RNA Polymerase. The amplified, biotin-labelled antisense RNA was purified, and its quality was assessed using the 2100 Bioanalyzer (Agilent) and the RNA 6000 Nano kit (Agilent). Antisense RNA (750 ng in 5 µl) was mixed with 10 µl of hybridization reagents and heated at 65 °C for 10 min. After cooling to room temperature, 15 µl hybridization solution was applied to a HumanHT-12 v4 chip (Illumina). The chip was incubated for about 18 h at 58 °C. After washing and staining with streptavidin–Cy3, the chip was scanned using a BeadArray Reader (Illumina). The scanning was done using a standard DirectHyb Gene Expression protocol with the following settings: Factor = 1, PMT = 587 and Filter = 100%. The raw data were extracted using BeadStudio software (Illumina) without normalization. Data were analysed using GeneSpring software with global normalization. Student's unpaired *t*-test was used to determine statistical significance.

31. Jones, C. T. *et al.* Real-time imaging of hepatitis C virus infection using a fluorescent cell-based reporter system. *Nature Biotechnol.* **28**, 167–171 (2010).
32. McGee, C. E. *et al.* Infection, dissemination, and transmission of a West Nile virus green fluorescent protein infectious clone by *Culex pipiens quinquefasciatus* mosquitoes. *Vector Borne Zoonotic Dis.* **10**, 267–274 (2009).
33. Tsetsarkin, K. *et al.* Infectious clones of chikungunya virus (La Reunion isolate) for vector competence studies. *Vector Borne Zoonotic Dis.* **6**, 325–337 (2006).
34. Suthar, M. S., Shabman, R., Madric, K., Lambeth, C. & Heise, M. T. Identification of adult mouse neurovirulence determinants of the Sindbis virus strain AR86. *J. Virol.* **79**, 4219–4228 (2005).
35. Bick, M. J. *et al.* Expression of the zinc-finger antiviral protein inhibits alphavirus replication. *J. Virol.* **77**, 11555–11562 (2003).
36. Lindenbach, B. D. *et al.* Complete replication of hepatitis C virus in cell culture. *Science* **309**, 623–626 (2005).
37. Lindenbach, B. D. & Rice, C. M. *trans*-Complementation of yellow fever virus NS1 reveals a role in early RNA replication. *J. Virol.* **71**, 9608–9617 (1997).
38. Petrakova, O. *et al.* Noncytopathic replication of Venezuelan equine encephalitis virus and eastern equine encephalitis virus replicons in mammalian cells. *J. Virol.* **79**, 7597–7608 (2005).
39. Zhang, Y. J. *et al.* Envelope-dependent, cyclophilin-independent effects of glycosaminoglycans on human immunodeficiency virus type 1 attachment and infection. *J. Virol.* **76**, 6332–6343 (2002).
40. Evans, M. J. *et al.* Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry. *Nature* **446**, 801–805 (2007).
41. Zennou, V. *et al.* HIV-1 genome nuclear import is mediated by a central DNA flap. *Cell* **101**, 173–185 (2000).
42. Jones, C. T., Murray, C. L., Eastman, D. K., Tassello, J. & Rice, C. M. Hepatitis C virus p7 and NS2 proteins are essential for production of infectious virus. *J. Virol.* **81**, 8374–8383 (2007).

LETTER

Digoxin and its derivatives suppress T$_H$17 cell differentiation by antagonizing RORγt activity

Jun R. Huh[1], Monica W. L. Leung[1]*, Pengxiang Huang[2]*, Daniel A. Ryan[3], Michael R. Krout[3], Raghu R. V. Malapaka[4], Jonathan Chow[1,5], Nicolas Manel[1]†, Maria Ciofani[1], Sangwon V. Kim[1], Adolfo Cuesta[1,5], Fabio R. Santori[1], Juan J. Lafaille[1], H. Eric Xu[4], David Y. Gin[3], Fraydoon Rastinejad[2] & Dan R. Littman[1,5]

CD4$^+$ T helper lymphocytes that express interleukin-17 (T$_H$17 cells) have critical roles in mouse models of autoimmunity, and there is mounting evidence that they also influence inflammatory processes in humans. Genome-wide association studies in humans have linked genes involved in T$_H$17 cell differentiation and function with susceptibility to Crohn's disease, rheumatoid arthritis and psoriasis[1–3]. Thus, the pathway towards differentiation of T$_H$17 cells and, perhaps, of related innate lymphoid cells with similar effector functions[4,5], is an attractive target for therapeutic applications. Mouse and human T$_H$17 cells are distinguished by expression of the retinoic acid receptor-related orphan nuclear receptor RORγt, which is required for induction of IL-17 transcription and for the manifestation of T$_H$17-dependent autoimmune disease in mice[6]. By performing a chemical screen with an insect cell-based reporter system, we identified the cardiac glycoside digoxin as a specific inhibitor of RORγt transcriptional activity. Digoxin inhibited murine T$_H$17 cell differentiation without affecting differentiation of other T cell lineages and was effective in delaying the onset and reducing the severity of autoimmune disease in mice. At high concentrations, digoxin is toxic for human cells, but non-toxic synthetic derivatives 20,22-dihydrodigoxin-21,23-diol and digoxin-21-salicylidene specifically inhibited induction of IL-17 in human CD4$^+$ T cells. Using these small-molecule compounds, we demonstrate that RORγt is important for the maintenance of IL-17 expression in mouse and human effector T cells. These data indicate that derivatives of digoxin can be used as chemical templates for the development of RORγt-targeted therapeutic agents that attenuate inflammatory lymphocyte function and autoimmune disease.

To identify small molecules that specifically inhibit transcriptional activity of RORγ and RORγt isoforms, we prepared *Drosophila* S2 cells stably expressing fusions of the GAL4 DNA-binding domain (DBD) and the ligand-binding domains (LBDs) of murine RORγ, RORα (mouse homologue of RORγ) and DHR3 (*Drosophila* orthologue for ROR family proteins), as well as the activation domain of the general transcriptional activator VP16. Induction of RORγ and the other fusion proteins led to robust expression of a firefly luciferase reporter (Supplementary Fig. 1a). Next, we investigated whether RORγ activity in the *Drosophila* system is dependent on a functional LBD and is ligand dependent. A single amino acid change in the putative ligand-binding pocket[7] of RORγ completely abrogated its function as a transcriptional activator despite a comparable level of protein expression both in S2 cells and in transgenic fly models (Supplementary Fig. 1b, c). In addition, *Drosophila* cells grown in serum-free media completely lacked RORγ activity, unless serum or cholesterol metabolites were supplemented into the cell culture (Supplementary Fig. 1d), indicating that yet-to-be-identified ligands are required for RORγ reporter

activity. These data justify utilization of the heterologous system to identify small molecules that modulate RORγ activity.

We next performed a chemical screen with 4,812 compounds and identified digoxin as a specific inhibitor for RORγ transcriptional activity (Fig. 1a). Digoxin inhibited RORγ (Fig. 1b and Supplementary Fig. 2a) with a half-maximum inhibitory concentration (IC$_{50}$) value of 1.98 μM. Inhibition of RORγ activity by digoxin was specific, as there was no effect on the transcriptional activity of VP16 or of the related nuclear hormone receptors RORα and DHR3 (Fig. 1c). Digoxin did not inhibit the activity of other nuclear hormone receptors, including *Caenorhabditis elegans* DAF-12, human androgen receptor and LXRα (Supplementary Fig. 2b, c). Digitoxin and β-acetyldigoxin also
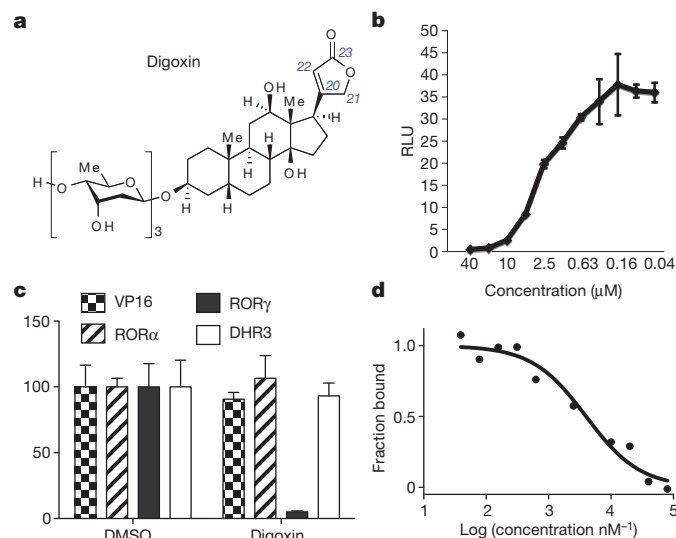

Figure 1 | Digoxin binds to RORγ and inhibits its transcriptional activity. a, Chemical structure of digoxin. **b**, Digoxin demonstrates dose-dependent inhibition of RORγ transcriptional activity in the *Drosophila* S2 cell luciferase reporter system. Ratio of firefly to *Renilla* luciferase activity is shown as relative luciferase unit (RLU) on the *y*-axis. **c**, Digoxin (10 μM) selectively inhibits RORγ-dependent transcriptional activity without affecting that of RORα, DHR3 or VP16. Percentages of relative luciferase units compared to DMSO-treated reporter cells are shown on the *y*-axis. Error bars indicate standard deviation. **d**, *In vitro* competition assay. Recombinant human RORγ LBD was loaded with fluorescently-labelled 25-hydroxycholesterol in the presence of the indicated concentrations of digoxin, and fluorescence polarization was measured.

[1]Molecular Pathogenesis Program, The Kimmel Center for Biology and Medicine of the Skirball Institute, New York University School of Medicine, New York, New York 10016, USA. [2]Sanford-Burnham Medical Research Institute at Lake Nona, 6400 Sanger Road, Orlando, Florida 32827, USA. [3]Molecular Pharmacology and Chemistry Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10065, USA. [4]Laboratory of Structural Sciences, Van Andel Research Institute, Grand Rapids, Michigan 49503, USA. [5]Howard Hughes Medical Institute, New York University School of Medicine, New York, New York 10016, USA. †Present address: Institut Curie, INSERM U932, 75005 Paris, France.
*These authors contributed equally to this work.

selectively inhibited RORγ (Supplementary Fig. 2d, e) with similar IC_{50} values. Next, we examined whether digoxin targets RORγ directly. 25-Hydroxycholesterol has been shown to bind to the RORγ LBD[8], and conjugation of fluorescein to this surrogate ligand did not affect its ability to bind to the human RORγ LBD (with a dissociation constant (K_d) of 109 nM). Addition of digoxin led to a dose-dependent decrease in fluorescence polarization values, demonstrating that digoxin can displace the sterol ligand with an IC_{50} of 4.1 μM (Fig. 1d). In addition, circular dichroism analysis showed that digoxin increased the thermal stability of the RORγ LBD, indicating that it interacts directly with RORγ (Supplementary Fig. 3a)[9]. Digoxigenin, the aglycone of digoxin, did not inhibit RORγt activity in *Drosophila* cells and did not bind to the RORγt LBD in the circular dichroism and competition assays (data not shown and Supplementary Fig. 3b). We further investigated whether digoxin binds inside the ligand-binding pocket of RORγ. We performed random mutagenesis on the LBD and screened 200 clones to identify those that were resistant to digoxin-mediated inhibition. Two clones with this property were identified and shared mutation of amino acid 290 (L290P/A494T and L290F/C318S). RORγ harbouring mutations at all three residues (RORγ/γt(triple)) exhibited much less sensitivity to digoxin, in spite of transcriptional activity similar to that of the wild-type molecule (Supplementary Fig. 3c, d). Two of the mutations mapped to the ligand-binding pocket (L290 and

C318) and one to helix 11 (A494)[8], consistent with digoxin binding inside the pocket.

When naive mouse CD4[+] T cells were cultured under T_H17 polarizing conditions (IL-6 and TGF-β), treatment with digoxin led to markedly reduced expression of IL-17a protein (Fig. 2a). Transcriptional upregulation of genes encoding the IL-23 receptor (IL-23R), IL-17a, IL-17f or IL-22 was also strongly inhibited (Supplementary Fig. 4a, b). Expression of RORγt-independent T_H17 signature genes, such as *Il21*, *Maf*, *Rora*, *Batf* and *Irf4*, was not affected by digoxin (Supplementary Fig. 4c, d). Reduction of T_H17 cell differentiation after treatment of wild-type cells with digoxin was similar to that observed upon targeted inactivation of *Rorc(γt)* (Fig. 2a). IL-23-induced T_H17 cell differentiation[10] was also inhibited in the presence of digoxin (Supplementary Fig. 4e). Importantly, digoxin had no effect on differentiation of naive CD4[+] T cells into other lineages, including T_H1, T_H2 and regulatory T cells (Supplementary Fig. 4f, g). Other cardiac glycosides with structures related to digoxin, including proscillaridin A, deslanoside, erysimoside, oleandrin, ouabain, ouabagenin, digitoxigenin, digoxigenin and lanatoside C, had no significant effect on RORγ transcriptional activity or T_H17 cell differentiation (Supplementary Fig. 5a, b).

To investigate if RORγt is the major target of digoxin or if another dominant cellular target exists, we performed gene expression profiling with total RNA samples isolated from dimethylsulphoxide (DMSO)- or digoxin-treated wild-type or RORγt-deficient cells cultured in T_H17 conditions. Treatment with digoxin resulted in changes in gene expression that were very similar to those observed in RORγt-deficient cells: two-way ANOVA analysis of differential gene expression revealed 67 genes that were significantly affected by the compound (DMSO versus digoxin) irrespective of genotype ($P < 0.05$) (Fig. 2b) and 323 that were affected by the genotype (wild type versus knockout) irrespective of compound treatment ($P < 0.05$) (Supplementary Fig. 6a). Ninety-four per cent of genes affected by digoxin treatment were similarly affected by RORγt deficiency. Importantly, no genes were significantly affected by the combination of gene inactivation and digoxin treatment. These results indicate that the effects of digoxin are predominantly mediated through RORγt. Induction of RORγt messenger RNA and protein expression was not affected by digoxin (Supplementary Fig. 4a, d). To rule out the possibility that digoxin
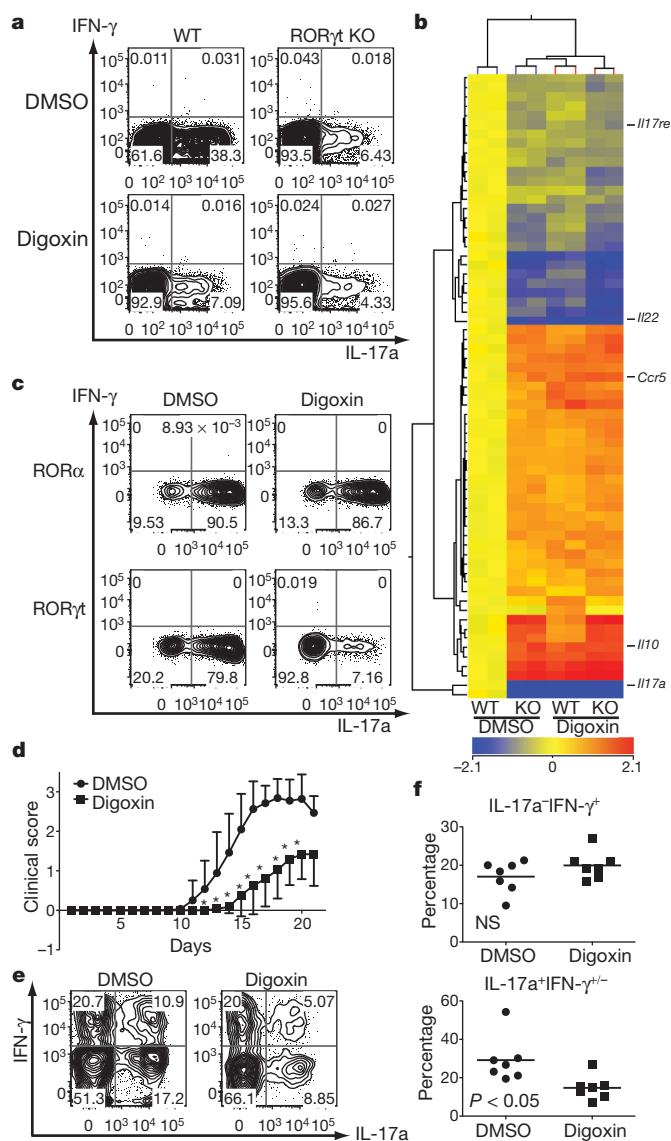

Figure 2 | Digoxin inhibits mouse T_H17 cell differentiation and ameliorates T_H17-mediated autoimmune disease. a, Flow cytometry of intracellular staining for IL-17a and IFN-γ in sorted naive T cell populations (from $Rorγt^{fl/fl}$ mice after transduction with control IRES–GFP (wild type (WT)) or CRE-IRES–GFP (knockout (KO)) retrovirus) activated and expanded in the presence of mouse T_H17 polarizing cytokines. DMSO or 10 μM digoxin was added at 6 h after viral transduction on day 1 and GFP-expressing cells were gated for analysis on day 5. **b**, Two-dimensional hierarchical clustering of the 67 genes (including redundant probe sets and genes of unknown function) identified to be significantly affected by two-way ANOVA analysis (DMSO versus digoxin treament, $P < 0.05$). Each row corresponds to a gene and each column corresponds to an experimental sample. **c**, FACS-sorted naive T cells were transduced with retroviral vectors encoding murine RORα–IRES–GFP or RORγt–IRES–GFP on day 1 (16 h after TCR stimulation) and GFP-expressing cells were gated for anyalsis on day 5. DMSO or 10 μM digoxin was added 6−8 h after transduction. **d**, EAE disease course in B6 wild-type mice that were injected i.p. with either DMSO or digoxin (40 μg per mouse) every day starting from day 2 after disease induction with myelin oligodendrocyte glycoprotein peptide 35–55 (MOG_{35-55}) in complete Freund's adjuvant (CFA). Shown is averaged curve shape from seven experiments (10 or 20 mice were used per trial). *$P < 0.05$. Error bars represent standard deviation. **e, f**, T_H1 and T_H17 cells in spinal cord of EAE mice treated with DMSO or digoxin. Lymphocytes were isolated on day 21 after disease induction. The cells were stimulated for 4 h with phorbol myristate acetate (PMA)/ionomycin and stained for surface markers and intracellular cytokines. Representative FACS plots (gated on CD45[+] CD11b[−] CD4[+] cells) from mice from each group are shown (**e**). T cells isolated from spinal cords of DMSO ($n = 7$) or digoxin treated mice ($n = 7$) were stained intracellularly for IFN-γ or IL-17a. Statistical analysis was by a two-tailed unpaired Student's t-test; NS, not significant and $P = 0.014$ (**f**).

blocks steps downstream of RORγt activity during T_H17 cell differentiation (for example, IL-17a production), we examined its effect on ectopic expression of RORγt or RORα in naive CD4[+] T cells. Both nuclear receptors were previously shown to be sufficient to induce IL-17a expression[11], presumably by binding to the same *cis*-acting elements[12]. Digoxin suppressed RORγ- and RORγt- but not RORα-mediated induction of IL-17a (Fig. 2c and Supplementary Fig. 6b), confirming that it acts selectively on RORγt in mouse T cells. However, digoxin (10 μM) failed to inhibit RORγt(triple)-mutant-mediated IL-17a production (Supplementary Fig. 3e). Digitoxin and β-acetyldigoxin also selectively inhibited RORγt-dependent T_H17 cell differentiation (Supplementary Fig. 6c). The aryl hydrocarbon receptor (AhR) is another ligand-dependent transcription factor that augments T_H17 responses[13]. Its activity was unaffected by digoxin, as addition of the AhR ligand 6-formylindolo[3,2-b]carbazole (FICZ) increased RORα-dependent IL-17a expression even in the presence of digoxin (Supplementary Fig. 6d). RORγt is predominantly found in the nucleus of T_H17 cells[14]. Digoxin treatment did not inhibit its nuclear localization in *Drosophila* cells or in *in vitro* differentiated T_H17 cells (Supplementary Fig. 7a, b). These data raise the question of how digoxin suppresses RORγt transcriptional activity. Chromatin immunoprecipitation-sequencing (ChIP-Seq) analysis with an anti-RORγt antibody (Supplementary Fig. 7c) was used for genome-wide identification of its transcriptional target sites in T_H17 cells (M. Ciofani and D.R.L., unpublished results). We evaluated the effect of digoxin on binding of RORγt to sites in two relevant loci, *Il17a/f* and *Il23r*. RORγt binding to these sites was substantially reduced on treatment with digoxin (Supplementary Fig. 7d), demonstrating one mode of its activity. *In vitro*, digoxin not only reduced the binding of RORγt onto its target, but also displaced SRC3-1b co-activator peptides (IC$_{50}$ of 1.8 μM) from the RORγt LBD and facilitated its interaction with co-repressor NCOR2 peptides (IC$_{50}$ of 3.9 μM) (Supplementary Figs 7e, 8a, b).

We next examined if digoxin can exert an anti-inflammatory effect in mice. We induced experimental autoimmune encephalomyelitis (EAE)—a T_H17-mediated autoimmune inflammatory disease—in C57BL/6 wild-type mice[15,16] in conjunction with intraperitoneal (i.p.) injections of digoxin or carrier each day from day 2. Digoxin treatment not only delayed the onset, but also reduced the severity of EAE progression (Fig. 2d). Also, the total number of mononuclear cells infiltrating the spinal cord was markedly reduced in mice treated with digoxin (Supplementary Fig. 9). Importantly, the percentage of IL-17-producing T cells infiltrating the spinal cord in digoxin-treated mice was reduced by more than 50% as compared to DMSO-treated mice, whereas that of IFN-γ-producing T_H1 cells was approximately the same (Fig. 2e, f). Administration of digoxigenin had no effect on the progression of EAE (data not shown), indicating that the cardiac glycoside activity[17] has no role in the observed amelioration of disease.

Digoxin, an inhibitor of the Na^+/K^+-ATPase, has long been used for treatment of congestive heart failure[17,18], but is toxic for human cells at concentrations (>300 nM)[19] well below those required for RORγt inhibition. Expression of the catalytic α1 subunit of murine Na^+/K^+-ATPase, which binds digoxin poorly, rendered human cells much less sensitive to digoxin-mediated cytotoxicity[20,21]. Thus, we ectopically expressed the α1 subunit of murine Na^+/K^+-ATPase in human cord blood CD4[+] T cells in the presence of cardiac glycosides. Lanatoside C (Supplementary Fig. 5a, b), which has inhibitory activity on the Na^+/K^+-ATPase similar to digoxin[17] but does not inhibit RORγt activity (data not shown), had no effect on IL-17a expression. However, digoxin suppressed IL-17a production (Supplementary Fig. 10a). Next, human T cells expressing the murine Na^+/K^+-ATPase were further transduced with lentivirus encoding human RORα, β, or γt, all of which are sufficient to induce IL-17 expression[22]. Digoxin inhibited only RORγt-mediated induction of IL-17a (Supplementary Fig. 10b), demonstrating its direct and selective activity on human RORγt.

Cardiac glycosides of the cardenolide class have three common structural motifs, namely a central steroidal core fused with a butenolide and

various sugars[23,24]. The glycans are dispensable, as digoxigenin still inhibits Na^+/K^+-ATPase[17]. 20,22-Dihydrodigoxin (Supplementary Fig. 2d), which was derived upon butenolide reduction of digoxin by the intestinal commensal *Eubacterium lentum*[25], has weak cardiac glycoside activity with much lower binding affinity than that of digoxin for Na^+/K^+-ATPase[17,26], yet it inhibited RORγ activity in the S2 reporter system (data not shown). As 20,22-dihydrodigoxin was still cytotoxic for human cells at 2.5 μM, digoxin was further modified by complete reduction of the butenolide to generate 20,22-dihydrodigoxin-21,23-diol (Dig(dhd)) (Fig. 3a). Dig(dhd) lacked cytotoxic activity on human cells at concentrations up to 40 μM, but it still possessed RORγ inhibitory activity and displaced the sterol ligand from the RORγ LBD (IC$_{50}$ of 12 μM) (Supplementary Fig. 11a). Additional derivatization of digoxin was achieved by aldol condensation of the butenolide with salicylaldehyde to produce digoxin-21-salicylidene (Dig(sal)) (Fig. 3a), which, similarly to digoxin, bound directly to RORγ in the circular dichroism assay (Supplementary Fig. 3a). These compounds selectively inhibited both mouse and human RORγt activities without affecting those of mouse RORα and human LXRβ (Supplementary Fig. 11b–d). Moreover, Dig(sal) treatment reduced the severity of EAE progression (Supplementary Fig. 11e). When tested on human CD4[+] T cells transduced with viruses encoding RORαd or RORγt, Dig(dhd) or Dig(sal) treatment selectively suppressed RORγt-mediated IL-17a induction (Supplementary Fig. 11f). Intriguingly, addition of either compound blocked T_H17 cell differentiation[22] (Fig. 3b) and induced reciprocal increases of IFN-γ or FOXP3 expression in T cells (Fig. 3b and Supplementary Fig. 11g), indicating that functional RORγt or its
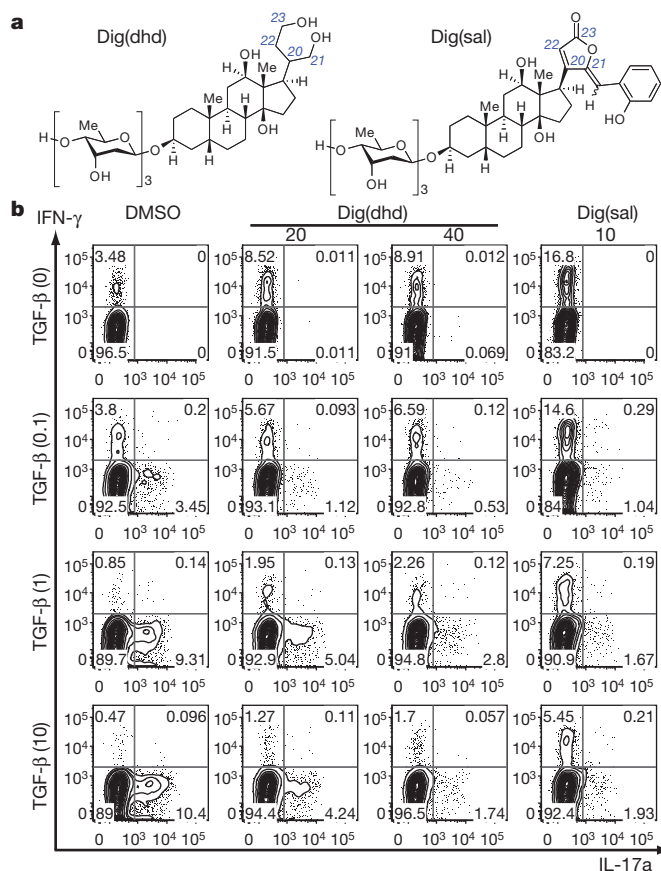

Figure 3 | Dig(dhd) and Dig(sal) inhibit human T_H17 cell differentiation.
a, Chemical structures of Dig(dhd) and Dig(sal). **b**, Flow cytometry of the production of IL-17a and IFN-γ by human naive cord blood T cells cultured for six days in the presence of IL-2, IL-23 and IL-1β, with various concentrations of TGF-β (ng ml^{-1}). DMSO, Dig(dhd) or Dig(sal) at indicated concentrations (μM) was added 16 h after cytokine addition.

downstream events may normally suppress development into other T cell lineages. Expression of another human T_H17 cell-associated surface marker, CCR6, was also reduced in Dig(dhd)-treated cells (Supplementary Fig. 11h).

We next investigated if digoxin can inhibit IL-17 production from pre-differentiated T_H17 cells. *In vitro* digoxin treatment of expanded mouse T_H17 cells derived from immunized mice inhibited both IL-23R (Fig. 4a) and IL-17a expression without affecting IFN-γ expression (Fig. 4b). We also purified GFP-positive T_H17 cells from MOG-immunized $Il23r^{gfp/+}$ mice[27] after 4-day *in vitro* culture with IL-23 and MOG peptide. More than 70% of the sorted GFP-positive cells expressed IL-17a (Supplementary Fig. 12a, day 0). GFP-positive cells were then treated with DMSO or digoxin for an additional 3 days. Digoxin treatment reduced IL-17a-expressing cells by more than 70% (Supplementary Fig. 12a, day 3), confirming that mouse T_H17 cells
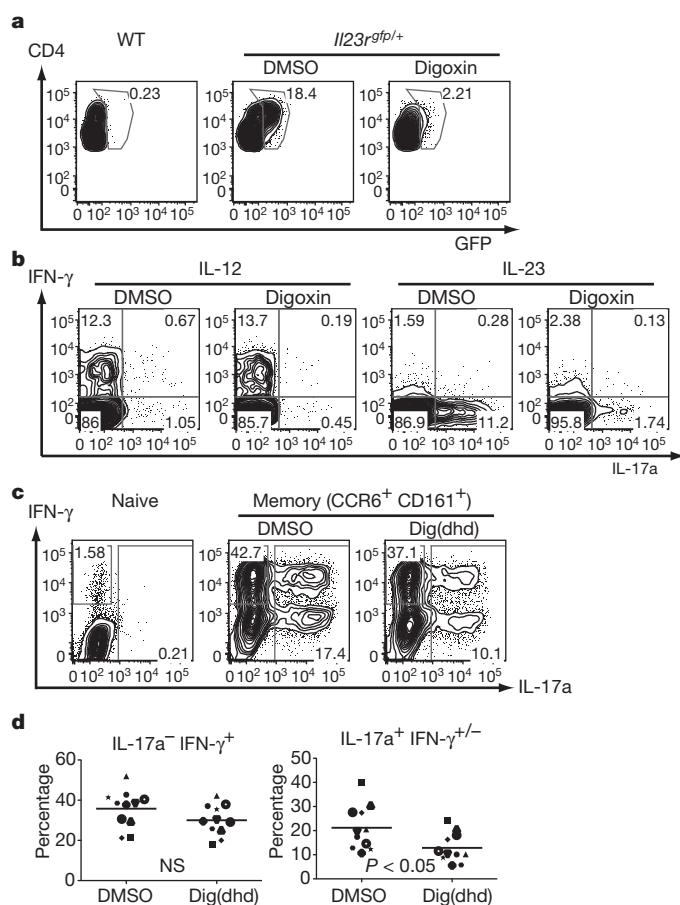

Figure 4 | RORγt activity is important for maintenance of mouse and human T_H17 cells. a, b, Flow cytometry of intracellular staining for IL-17a and IFN-γ by CD4$^+$ T cells. Mononuclear cells were collected from draining lymph nodes of wild-type or IL-23R–GFP knock-in heterozygous mice 7 days after MOG$_{35-55}$/CFA injection. Cells were cultured for four more days with MOG$_{35-55}$ peptide and exogenous IL-23 or IL-12, in the presence of DMSO or 10 μM digoxin. Without pre-immunization, addition of IL-23 and MOG$_{35-55}$ peptide to culture did not lead to *de novo* T_H17 cell differentiation (data not shown). Digoxin treatment suppressed expansion of *in vivo* differentiated T_H17 cells, assayed by IL-23R reporter GFP expression (**a**) or by IL-17a production (**b**). **c, d,** Human naive (CD45RA$^+$ CD3$^+$ CD4$^+$) or memory (CD45RO$^+$ CD45RA$^-$ CD3$^+$ CD4$^+$ CCR6$^+$ CD161$^+$) cells were purified from healthy donor peripheral blood samples and were cultured in the presence of IL-1β, IL-23 and IL-2 for 6 days with or without 40 μM Dig(dhd). Representative FACS plots from one donor are shown (**c**). Each symbol indicates a separate donor ($n = 11$). Statistical analysis was by a two-tailed unpaired Student's *t*-test; IL-17a$^-$ IFN-γ^+, NS, and IL-17a$^+$ IFN-$\gamma^{+/-}$, $P = 0.02$ (**d**).

generated *in vivo* and expanded *in vitro* require continuous RORγt activity to maintain their identity. To test if digoxin suppresses the activity of pre-existing T_H17 cells *in vivo*, we transferred IFN-γ-deficient, MBP-specific T_H17 cells into lymphopenic RAG2-deficient mice and assessed EAE manifestation after daily administration of digoxin. Because the transferred cells lack IFN-γ, the EAE phenotype observed in these mice is entirely attributable to the function of T_H17 cells. Digoxin treatment from day 2 delayed onset and reduced severity of T_H17 cell transfer-mediated EAE, which further confirms a requirement for continuous RORγt activity in T_H17 cells (Supplementary Fig. 12b). We then examined if RORγt activity is also important for sustained expression of IL-17a in human CD4$^+$ T cells. Human memory T_H17 cells were purified from peripheral blood samples and enriched by *in vitro* culture. Naive CD4$^+$ T cells cultured in the same cytokine conditions did not produce IL-17a (Fig. 4c, left plot). Dig(dhd) treatment led to a 40–50% reduction of IL-17a-expressing cells with little effect on IFN-γ-expressing cells (Fig. 4c, d). These data demonstrate that human RORγt activity has an important role in maintaining the human T_H17 cell population.

T cells and innate lymphocytes that produce IL-17a and IL-22 are recognized as having key roles in maintenance of barrier function at mucosal surfaces and also in the pathophysiology of autoimmune disease. All such cells, which include T_H17 cells, other TCRαβ cells, TCRγδ cells, lymphoid tissue inducer cells and NK-like cells (also referred to as NK22 cells) share in the property of requiring expression of RORγt for their differentiation. Abrogation of RORγt expression results in marked reduction or complete depletion of these cell types and in resistance to T_H17-mediated autoimmune disease in mouse models[4,6,28]. Therefore, RORγt antagonists digoxin, Dig(dhd) and Dig(sal) may serve as good chemical templates for the development of potent therapeutic compounds to treat various diseases associated with inflammatory lymphocyte dysfunction.

The digitalis-like compounds were originally identified in plants. A body of evidence indicates the presence of endogenous digitalis-like compounds in mammals including humans[18,29]. Identification of digoxin as a RORγt antagonist suggests that related molecules in mammals may modulate RORγ- and RORγt-mediated functions. However, it would be derivative compounds with better IC$_{50}$ values that would have such roles. In light of recent findings of the roles of microbiota in the generation of T_H17 cells in the small intestine[30], it is interesting that *Eubacterium lentum*—another commensal bacterium—has the capacity to metabolize digoxin into dihydrodigoxin. The possibility of the existence of endogenous digitalis-like compounds in host organisms and of their modification by microbes may present further opportunities for modulating the function of RORγt and T_H17 cell differentiation.

METHODS SUMMARY

Chemical screen. Ten thousand *Drosophila* stable S2 cells with genomic integration of the Cu^{2+}-inducible G4DBD-mouse RORγ construct were transfected with 5 ng of pUAST firefly luciferase and 7 ng of Pol III *Renilla* luciferase and dispensed into white-bottom tissue-culture 384-well plates (Corning). Two days later, small compounds (total 4,812 compounds from the ICCB chemical libraries, including Bioactives and Prestwick collections) were added and, after 6 h, Cu^{2+} was added to the wells (700 μM). The following morning, Stop-glo luciferase substrates (Promega) were used to measure luciferase activity. Initial hits including digoxin were tested against three different control S2 reporter cell lines.

Cell culture. Mouse and human CD4$^+$ T cell culture and viral transduction were performed as described previously[6,22], unless indicated otherwise in the text.

Identification of non-toxic digoxin derivatives. Various digoxin derivatives were synthesized and first tested for toxicity on human embryonic kidney 293T cells at various concentrations. Compounds exhibiting reduced toxicity compared to digoxin were further tested for their RORγ inhibitory activities with the insect cell reporter lines.

General. All DNA constructs were generated by PCR-based methodology and confirmed by sequencing. Retroviral production and transduction, EAE experiments and gene chip analysis were performed as described previously[6]. IL-17a, IFN-γ, IL-4, FOXP3 and CCR6 protein expression was examined by intracellular or surface staining according to the manufacturer's protocol.

1. Duerr, R. H. *et al.* A genome-wide association study identifies *IL23R* as an inflammatory bowel disease gene. *Science* **314**, 1461–1463 (2006).
2. Nair, R. P. *et al.* Genome-wide scan reveals association of psoriasis with IL-23 and NF-κB pathways. *Nature Genet.* **41**, 199–204 (2009).
3. Stahl, E. A. *et al.* Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nature Genet.* **42**, 508–514 (2010).
4. Buonocore, S. *et al.* Innate lymphoid cells drive interleukin-23-dependent innate intestinal pathology. *Nature* **464**, 1371–1375 (2010).
5. Colonna, M. Interleukin-22-producing natural killer cells and lymphoid tissue inducer-like cells in mucosal immunity. *Immunity* **31**, 15–23 (2009).
6. Ivanov, I. I. *et al.* The orphan nuclear receptor RORγt directs the differentiation program of proinflammatory IL-17$^+$ T helper cells. *Cell* **126**, 1121–1133 (2006).
7. Stehlin, C. *et al.* X-ray structure of the orphan nuclear receptor RORβ ligand-binding domain in the active conformation. *EMBO J.* **20**, 5822–5831 (2001).
8. Jin, L. *et al.* Structural basis for hydroxycholesterols as natural ligands of orphan nuclear receptor RORγ. *Mol. Endocrinol.* **24**, 923–929 (2010).
9. Raghuram, S. *et al.* Identification of heme as the ligand for the orphan nuclear receptors REV-ERBα and REV-ERBβ. *Nature Struct. Mol. Biol.* **14**, 1207–1213 (2007).
10. Ghoreschi, K. *et al.* Generation of pathogenic T$_H$17 cells in the absence of TGF-β signalling. *Nature* **467**, 967–971 (2010).
11. Yang, X. O. *et al.* T helper 17 lineage differentiation is programmed by orphan nuclear receptors RORα and RORγ. *Immunity* **28**, 29–39 (2008).
12. Sato, T. K. *et al.* A functional genomics strategy reveals Rora as a component of the mammalian circadian clock. *Neuron* **43**, 527–537 (2004).
13. Veldhoen, M. *et al.* The aryl hydrocarbon receptor links T$_H$17-cell-mediated autoimmunity to environmental toxins. *Nature* **453**, 106–109 (2008).
14. Zhou, L. *et al.* TGF-β-induced Foxp3 inhibits T$_H$17 cell differentiation by antagonizing RORγt function. *Nature* **453**, 236–240 (2008).
15. Langrish, C. L. *et al.* IL-23 drives a pathogenic T cell population that induces autoimmune inflammation. *J. Exp. Med.* **201**, 233–240 (2005).
16. Cua, D. J. *et al.* Interleukin-23 rather than interleukin-12 is the critical cytokine for autoimmune inflammation of the brain. *Nature* **421**, 744–748 (2003).
17. Paula, S., Tabet, M. R. & Ball, W. J. Jr. Interactions between cardiac glycosides and sodium/potassium-ATPase: three-dimensional structure–activity relationship models for ligand binding to the E_2-P_i form of the enzyme versus activity inhibition. *Biochemistry* **44**, 498–510 (2005).
18. Nesher, M., Shpolansky, U., Rosen, H. & Lichtstein, D. The digitalis-like steroid hormones: new mechanisms of action and biological significance. *Life Sci.* **80**, 2093–2107 (2007).
19. Johansson, S. *et al.* Cytotoxicity of digitoxin and related cardiac glycosides in human tumor cells. *Anticancer Drugs* **12**, 475–483 (2001).
20. Price, E. M. & Lingrel, J. B. Structure-function relationships in the Na,K-ATPase α subunit: site-directed mutagenesis of glutamine-111 to arginine and asparagine-122 to aspartic acid generates a ouabain-resistant enzyme. *Biochemistry* **27**, 8400–8408 (1988).
21. Zavareh, R. B. *et al.* Inhibition of the sodium/potassium ATPase impairs N-glycan expression and function. *Cancer Res.* **68**, 6688–6697 (2008).
22. Manel, N., Unutmaz, D. & Littman, D. R. The differentiation of human T$_H$-17 cells requires transforming growth factor-β and induction of the nuclear receptor RORγt. *Nature Immunol.* **9**, 641–649 (2008).
23. Prassas, I. & Diamandis, E. P. Novel therapeutic applications of cardiac glycosides. *Nature Rev. Drug Discov.* **7**, 926–935 (2008).
24. Mijatovic, T. *et al.* Cardiotonic steroids on the road to anti-cancer therapy. *Biochim. Biophys. Acta* **1776**, 32–57 (2007).
25. Robertson, L. W., Chandrasekaran, A., Reuning, R. H., Hui, J. & Rawal, B. D. Reduction of digoxin to 20R-dihydrodigoxin by cultures of *Eubacterium lentum*. *Appl. Environ. Microbiol.* **51**, 1300–1303 (1986).
26. Lindenbaum, J., Rund, D. G., Butler, V. P. Jr, Tse-Eng, D. & Saha, J. R. Inactivation of digoxin by the gut flora: reversal by antibiotic therapy. *N. Engl. J. Med.* **305**, 789–794 (1981).
27. Awasthi, A. *et al.* Cutting edge: IL-23 receptor gfp reporter mice reveal distinct populations of IL-17-producing cells. *J. Immunol.* **182**, 5904–5908 (2009).
28. Luci, C. *et al.* Influence of the transcription factor RORγt on the development of NKp46$^+$ cell populations in gut and skin. *Nature Immunol.* **10**, 75–82 (2009).
29. Bagrov, A. Y. & Shapiro, J. I. Endogenous digitalis: pathophysiologic roles and therapeutic applications. *Nat. Clin. Pract. Nephrol.* **4**, 378–392 (2008).
30. Ivanov, I. I. *et al.* Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell* **139**, 485–498 (2009).

LETTER

Suppression of T$_H$17 differentiation and autoimmunity by a synthetic ROR ligand

Laura A. Solt[1], Naresh Kumar[1,2], Philippe Nuhant[3], Yongjun Wang[1], Janelle L. Lauer[1], Jin Liu[1], Monica A. Istrate[1,2], Theodore M. Kamenecka[4], William R. Roush[3], Dušica Vidović[2,5], Stephan C. Schürer[2,5], Jihong Xu[6], Gail Wagoner[6], Paul D. Drew[6], Patrick R. Griffin[1,2,4] & Thomas P. Burris[1]

T-helper cells that produce interleukin-17 (T$_H$17 cells) are a recently identified CD4$^+$ T-cell subset with characterized pathological roles in autoimmune diseases[1–3]. The nuclear receptors retinoic-acid-receptor-related orphan receptors α and γt (RORα and RORγt, respectively) have indispensible roles in the development of this cell type[4–7]. Here we present SR1001, a high-affinity synthetic ligand—the first in a new class of compound—that is specific to both RORα and RORγt and which inhibits T$_H$17 cell differentiation and function. SR1001 binds specifically to the ligand-binding domains of RORα and RORγt, inducing a conformational change within the ligand-binding domain that encompasses the repositioning of helix 12 and leads to diminished affinity for co-activators and increased affinity for co-repressors, resulting in suppression of the receptors' transcriptional activity. SR1001 inhibited the development of murine T$_H$17 cells, as demonstrated by inhibition of interleukin-17A gene expression and protein production. Furthermore, SR1001 inhibited the expression of cytokines when added to differentiated murine or human T$_H$17 cells. Finally, SR1001 effectively suppressed the clinical severity of autoimmune disease in mice. Our data demonstrate the feasibility of targeting the orphan receptors RORα and RORγt to inhibit specifically T$_H$17 cell differentiation and function, and indicate that this novel class of compound has potential utility in the treatment of autoimmune diseases.

T$_H$17 cells are crucial effector cells implicated in the pathology of numerous autoimmune diseases, including multiple sclerosis, rheumatoid arthritis, inflammatory bowel disease and systemic lupus erythematosis. These cells produce a number of cytokines, including interleukin-17 (IL-17), that are known to enhance inflammatory processes[1–3]. The discovery of T$_H$17 cells as critical mediators of autoimmune disorders provides a unique opportunity to develop focused therapeutics that act by inhibiting the function of these cells. An essential role for the nuclear receptors RORα and RORγt has been established in the development of T$_H$17 cells. Both of these nuclear receptors are required for the full differentiation of naive CD4$^+$ T cells into T$_H$17 cells[4–7].

Members of the nuclear receptor superfamily are ligand-dependent transcription factors. A number of drugs used in the clinic have been developed that target several nuclear receptor superfamily members. Therefore, an attractive strategy for the development of novel therapeutics to treat T$_H$17-mediated autoimmune disorders is to target RORα and RORγt with synthetic ligands that inhibit their activity, resulting in decreased T$_H$17 cell differentiation and/or function. However, RORs are generally characterized as orphan receptors with no well-characterized ligands, thus it is unclear whether this approach is viable.

We recently characterized the benzenesulphonamide liver X receptor (LXR) agonist T0901317 as a promiscuous ligand that modulates the activity of several nuclear receptors including RORα and RORγ[8]. Although T0901317 is a very potent and efficacious agonist of LXR, it also acts as an inverse agonist of RORα and RORγ by suppressing their basal transcriptional activity[8]. Using the T0901317 scaffold as a lead compound we developed a derivative, SR1001 (Fig. 1a and Supplementary Fig. 1), that was devoid of all LXR activity yet retained its ability to suppress the activity of RORα and RORγ. We found that SR1001
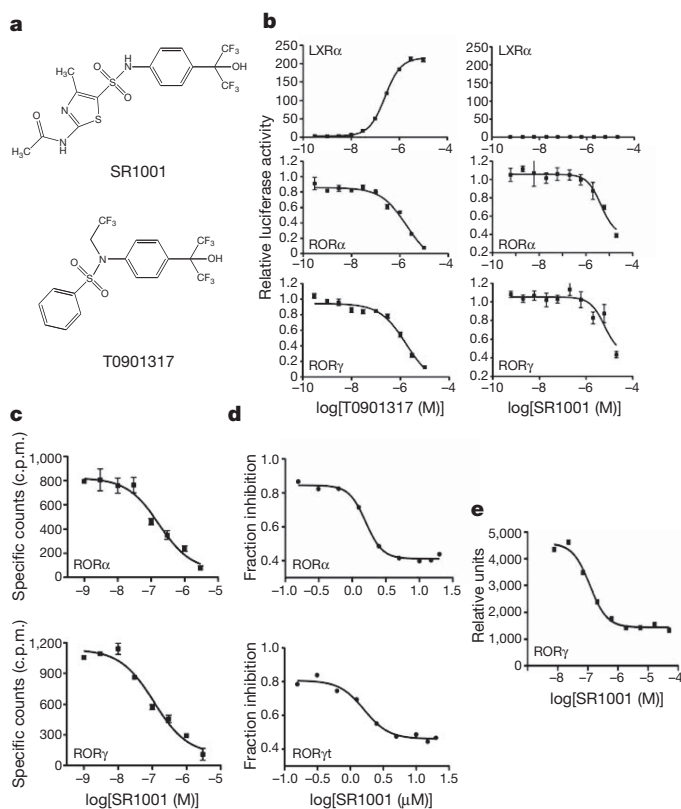
Figure 1 | SR1001 is a selective RORα and RORγ inverse agonist.
a, Structure of SR1001 and T0901317. b, GAL4–LXRα, GAL4–RORα and GAL4–RORγ co-transfection assays in HEK293 cells comparing T0901317 to SR1001 ($n = 8$). c, Competition radioligand binding assays illustrating the direct binding of SR1001 to the LBD of RORα and RORγ relative to [^{3}H]25-hydroxycholesterol ($n = 4$). c.p.m., counts per minute. d, SR1001 dose-dependently inhibits an *Il17* promoter-driven luciferase construct in the presence of RORα or RORγt in HEK293 cells. Results are normalized to vehicle (DMSO) control ($n = 4$). e, Alpha screen assay indicating that SR1001 dose-dependently inhibits the recruitment of a TRAP220 NR box 2 peptide to the LBD of RORγ ($n = 3$). All error bars denote s.e.m.

[1]Department of Molecular Therapeutics, The Scripps Research Institute, Jupiter, Florida 33458, USA. [2]The Scripps Research Institute Molecular Screening Center, The Scripps Research Institute, Jupiter, Florida 33458, USA. [3]Department of Chemistry, The Scripps Research Institute, Jupiter, Florida 33458, USA. [4]The Translational Research Institute, The Scripps Research Institute, Jupiter, Florida 33458, USA. [5]Center for Computational Science University of Miami, Miami, Florida 33136, USA. [6]Department of Neurobiology and Developmental Sciences, University of Arkansas for Medical Sciences, Little Rock, Arkansas 72205, USA.

repressed both GAL4–RORα and GAL4–RORγ transcriptional activity in a dose-dependent manner (Fig. 1b), but demonstrated no effect on LXRα activity (Fig. 1b). We assessed the specificity of SR1001 in a panel comprising all 48 human nuclear receptors in a cell-based co-transfection assay[8] and did not observe activity on receptors other than RORα or RORγ (data not shown). We examined the direct binding of SR1001 to RORα and RORγ using competitive radio-ligand binding assays. SR1001 dose-dependently displaced [^{3}H]25-hydroxycholesterol binding to RORα and RORγ (inhibition constant (K_i) = 172 and 111 nM, respectively) (Fig. 1c) but demonstrated no binding to RORβ (data not shown).

We examined whether SR1001 would affect RORα- and RORγ-dependent regulation of an $Il17$ promoter-driven luciferase reporter[9]. HEK293 cells were transfected with the $Il17$ reporter and either full-length RORα or RORγ and treated with SR1001 or vehicle. As shown in Fig. 1d, SR1001 dose-dependently suppressed the $Il17$ promoter-driven activity by each of the receptors. Because SR1001 bound RORα and RORγ, resulting in suppression of each receptor's transcriptional activity, we expected that SR1001 would inhibit co-activator binding to the receptors. SR1001 reduced the interaction of a co-activator TRAP220 nuclear receptor box 2 peptide with RORγ in a dose-dependent manner (Fig. 1e) (half-maximum inhibitory concentration (IC$_{50}$) value of ~117 nM). Collectively, these data demonstrate that SR1001 functions as an inverse agonist ligand of RORα and RORγ.

Next, we determined whether SR1001 affected endogenous $Il17a$ gene expression. The EL4 murine tumour cell line constitutively expresses RORα (encoded by $Rora$), RORγt ($Rorc$) and IL-17A ($Il17a$)[10]. EL4 cells were treated with either control short interfering RNA (siRNA) or a mixture of RORα/γ siRNA followed by treatment with either vehicle or SR1001. Reduction in the expression of RORα and RORγt significantly reduced the expression of $Il17a$ messenger RNA as measured by quantitative polymerase chain reaction (PCR) (Fig. 2a). More importantly, treatment of cells with SR1001 suppressed $Il17a$ mRNA expression whereas treatment of RORα/γ-depleted cells displayed a significantly blunted response, indicating that SR1001 suppression of $Il17a$ mRNA expression is RORα/RORγ dependent (Fig. 2a). Furthermore, SR1001 suppressed the expression of the RORα and RORγ target gene $G6PC$ in HepG2 cells, a human hepatocellular carcinoma cell line, providing further proof that the effect of SR1001 is mediated by RORα and RORγ (Supplementary Fig. 2)[11,12].

We hypothesized that SR1001 would inhibit binding of the co-activator SRC2 to either RORα or RORγ when these receptors are occupying the $Il17$ promoter. We performed a sequential chromatin immunoprecipitation assay (ChIP-reChIP) assessing the relative amount of SRC2 associated with either RORα or RORγ resident at the $Il17$ promoter in EL4 cells. SR1001 suppressed the ability of SRC2 to bind to RORα and RORγ at the $Il17$ promoter and increased the recruitment of the co-repressor NCoR (Fig. 2b, lanes 3 and 4 and Fig. 2c, lanes 3 and 4). Thus, SR1001 suppresses $Il17a$ expression by directly inhibiting co-activator binding and promoting the recruitment of co-repressors to RORα and RORγ.

To understand how ligand mediates the transcriptional activation of RORγ, we performed comprehensive differential hydrogen–deuterium exchange mass spectrometry (HDX) analysis of the RORγ ligand-binding domain (LBD) in the presence and absence of SR1001. This approach provides a measure of the localized ligand-induced perturbation in the conformational ensemble of the receptor. HDX kinetics of peptic peptides derived from the RORγ LBD were measured and the average difference in the percentage of incorporated deuterium between apo RORγ LBD and SR1001-bound RORγ LBD is presented in Supplementary Fig. 3. A negative value represents an increase in protection to exchange (more stable, less dynamic) in that region of the LBD when bound to ligand as compared to apo, whereas a positive value represents a decrease in protection to exchange (less stable, more dynamic). HDX kinetics are sensitive to hydrogen bond networks, and perturbations in these networks upon ligand binding can be determined
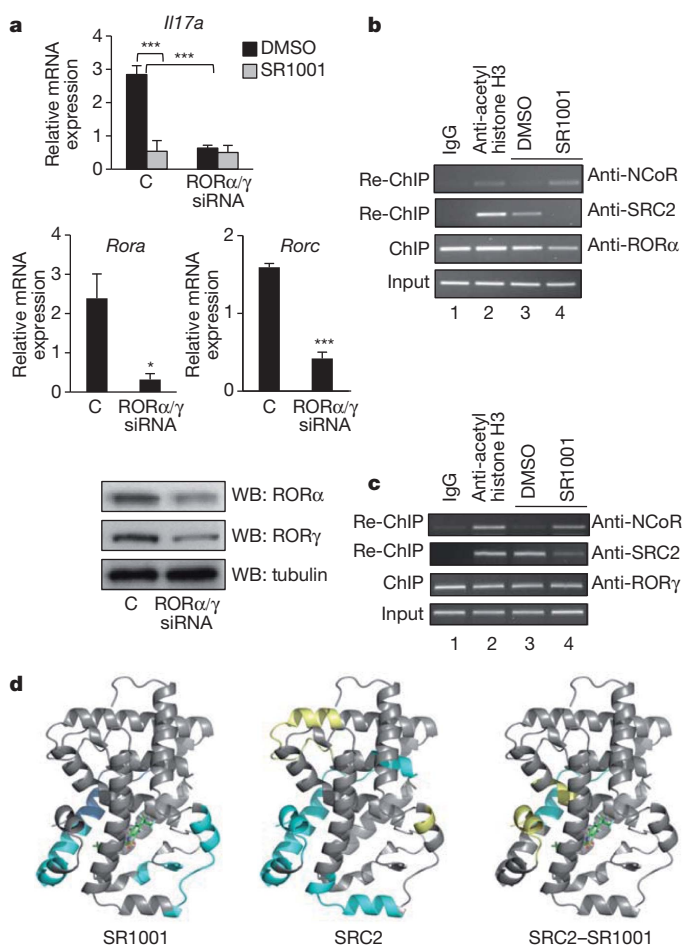

Figure 2 | SR1001 modulates the expression of ROR target genes by decreasing co-activator recruitment. a, $Il17a$, $Rora$ and $Rorc$ mRNA expression in EL4 cells treated with control (C) or mouse RORα/γ siRNA, and vehicle (DMSO) or SR1001 (10 μM, 24 h) ($n = 3$). Protein expression of RORα and RORγ is shown. *, $P < 0.05$; ***, $P < 0.005$. Error bars denote s.e.m. **b, c**, ChIP-reChIP assay in EL4 cells showing that SR1001 reduces RORα-dependent (**b**) and RORγ-dependent (**c**) recruitment of SRC2 and promotes the recruitment of NCoR to the $Il17$ promoter. **d**, Illustration of the HDX kinetics of peptic peptides derived from the RORγ LBD. Cyan indicates an increase in protection to exchange; yellow represents a decrease in protection to exchange.

using differential HDX. The differential HDX induced by SR1001 binding to RORγ correlates with the co-crystal structure of RORγ in complex with the sterol ligand 25-hydroxycholesterol (25-OHC; Protein Data Bank (PDB) 3LOL)[13] (Supplementary Fig. 4). In the RORγ–25-OHC structure, the C25 hydroxyl tail is oriented towards helix 11 (H11) and the A ring towards H1 and H2. As can be inferred from the RORγ–25-OHC structure (PDB 3LOL), the hydroxyl group at the C1 position of 25-OHC is hydrogen bonded to Gln 286 (H1) and the 25-hydroxyl is hydrogen bonded to His 479 (H11). The regions within the RORγ LBD that show increased protection to exchange upon binding of SR1001 include portions of H1 and H11. To highlight this, the HDX data in Supplementary Fig. 3 are represented graphically by overlay onto the structure of RORγ–25-OHC (PDB 3LOL) with SR1001 docked (Fig. 2d). Consistent with the differential HDX data, docking of SR1001 to RORγ–25-OHC (PDB 3LOL) indicates a similar binding mode for SR1001 to RORγ (Supplementary Fig. 4).

To examine the role of SR1001 in the modulation of this interaction between SRC2 and RORγ, we performed differential HDX on RORγ LBD in the presence and absence of the receptor interaction domain (RID) of SRC2 (Fig. 2d and Supplementary Fig. 3), which contains three nuclear receptor boxes (~18 kDa). Several regions of the LBD demonstrate reduced HDX kinetics in the presence of SRC2 RID,
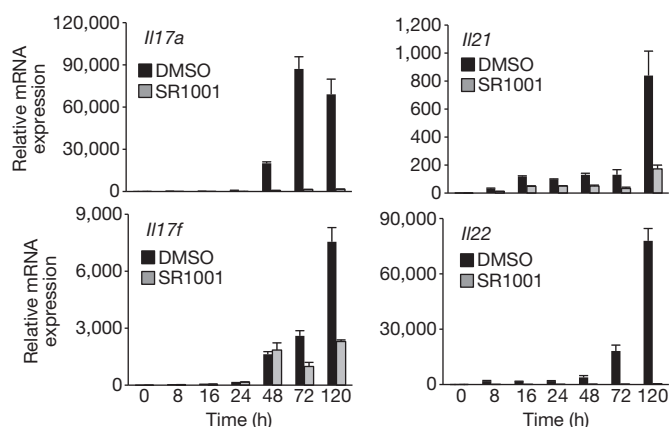
Figure 3 | SR1001 inhibits the expression of cytokines expressed by T$_H$17 cells. *Il17a*, *Il17f*, *Il21* and *Il22* mRNA expression in splenocytes differentiated under T$_H$17 polarizing conditions in the presence of vehicle (DMSO) or SR1001 (5 μM) for 5 days. Messenger RNA expression levels are normalized to *Gapdh* (*n* = 3). Error bars denote s.e.m.

indicating an interaction between the two proteins. One region that is stabilized is H12, containing the AF2 domain of the receptor, which has been shown to be important for nuclear receptor interaction with co-activators. Furthermore, differential HDX analysis of the RORγ–SRC2 complex in the presence and absence of SR1001 clearly demonstrates that ligand disrupts the receptor's interaction with SRC2 RID (Fig. 2d). These data provide strong mechanistic insight into how inverse agonists such as SR1001 repress transcriptional output of RORγ target genes.

Because RORα and RORγt activity is required for optimal T$_H$17 cell development[4], we explored whether SR1001 would inhibit T$_H$17 cell differentiation. Splenocytes were cultured under T$_H$17 polarizing conditions (TGF-β and IL-6) with SR1001 or vehicle control for 5 days. The combination of TGF-β and IL-6 increased the mRNA expression of *Il17a*, *Il17f*, *Il21* and *Il22* in vehicle-treated cells, whereas SR1001-treated cells failed to significantly upregulate these cytokines (Fig. 3). Propidium iodide staining indicated that SR1001 was not toxic and did not induce cell death (Supplementary Fig. 5). T$_H$17 cells and inducible T-regulatory cells (inducible T$_{reg}$) are both dependent on TGF-β for their differentiation. We evaluated whether expression of the T$_{reg}$-specific transcription factor Foxp3 was affected by SR1001 treatment.

Similar to vehicle control, *Foxp3* mRNA expression was unaffected by SR1001 treatment, indicating that inhibition of T$_H$17 cell differentiation by SR1001 did not drive the cells into an inducible T$_{reg}$ phenotype (Supplementary Fig. 6)[9]. Furthermore, suppression of T$_H$17 cell development with SR1001 treatment did not drive the splenocyte cultures into any of the other T-helper lineages, T$_H$1 or T$_H$2, as indicated by the decrease in *Tbx21* (T-bet) and *Gata3* mRNA expression, respectively (Supplementary Fig. 6).

Finally, we explored whether SR1001 would inhibit IL-17 protein production and secretion. Splenocytes were cultured under T$_H$17 polarizing conditions and analysed for IL-17 expression by intracellular flow cytometry. Treatment with SR1001 inhibited the expression of IL-17 from CD4$^+$ T cells at day 4, 5 and 6 (Fig. 4a). Similar to splenocyte cultures, intracellular flow cytometry demonstrated that SR1001 significantly repressed IL-17 expression in purified differentiated murine CD4$^+$ T cells (CD4$^+$CD25$^-$CD62L^{hi}CD44^{lo}) (Fig. 4b). Next we assessed the effect of SR1001 on IL-17 secretion from splenocyte cultures by enzyme-linked immunosorbent assay (ELISA). Treatment with SR1001 inhibited IL-17 secretion over a 3-day time course, when SR1001 was added at either the initiation of T$_H$17 cell differentiation (initiation) or 48 h after initiation of differentiation (post) (Fig. 4c). SR1001 was also effective at inhibiting intracellular IL-17 expression in human peripheral blood mononuclear cells (PBMCs) (Fig. 4d). Finally, we examined the effects of SR1001 on other T-helper cell lineages. Differentiation of T$_H$1, T$_H$2 and inducible T$_{reg}$ cells was unaffected by SR1001 treatment, as similar amounts of interferon (IFN)-γ, IL-4, or Foxp3, respectively, were expressed compared to vehicle controls, indicating that SR1001 specifically targets T$_H$17 cells (Supplementary Fig. 7).

Given that RORα and RORγt are required for the development of T$_H$17-mediated autoimmune diseases[4,5] and that SR1001 inhibits the activity of both of these receptors leading to suppressed T$_H$17 cell development *in vitro*, we evaluated the effects of SR1001 treatment in an animal model of multiple sclerosis, experimental autoimmune encephalomyelitis (EAE), a well characterized model of T$_H$17-cell-mediated autoimmune disease[14,15]. After myelin oligodendrocyte glycoprotein (MOG$_{35-55}$) immunization at day 0, mice were treated with 25 mg kg^{-1} of SR1001 twice per day intraperitoneally for the duration of the study. As shown in Fig. 4e, SR1001 treatment delayed the onset and clinical severity of EAE. Further analysis of spinal cords collected from mice at day 18 after immunization revealed that SR1001 repressed *Il17a* mRNA expression by ~60%, as well as reduced *Il21*
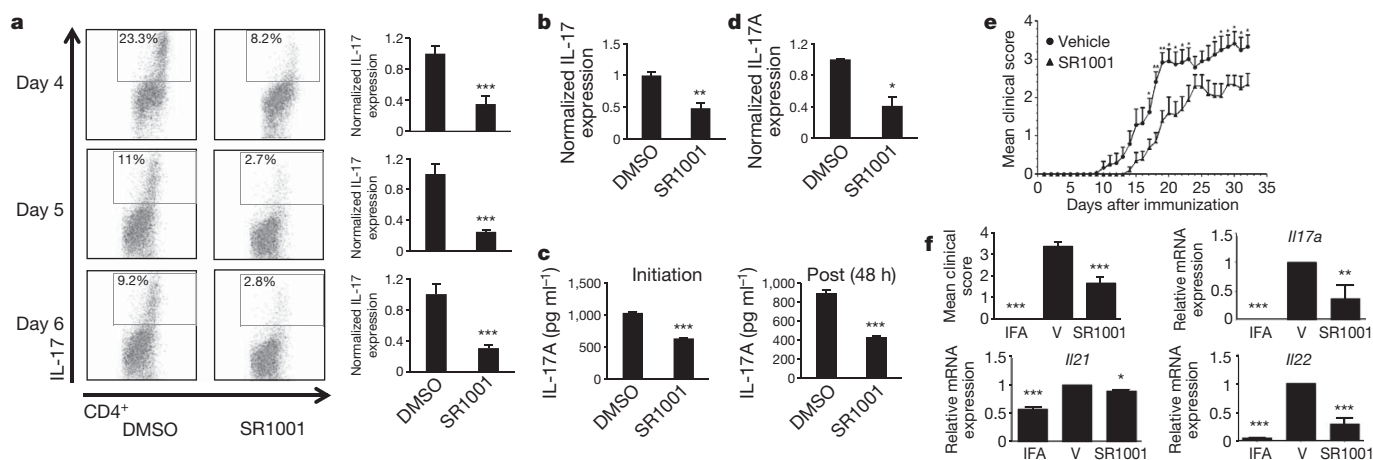
Figure 4 | SR1001 inhibits T$_H$17 cell development and IL-17A secretion.
a, IL-17 expression in splenocytes cultured under T$_H$17 polarizing conditions with vehicle control (DMSO) or SR1001 (5 μM). Graphs represent the average percentage of IL-17A expressing cells normalized to vehicle control (*n* = 3). **b**, IL-17 expression in differentiated purified naive murine CD4$^+$ T cells normalized to vehicle control (*n* = 3). **c**, IL-17A secretion from splenocytes cultured under T$_H$17 polarizing conditions with SR1001 (5 μM) for 3 days (*n* = 3). **d**, Intracellular IL-17A expression in human PBMCs cultured for 24 h with vehicle or SR1001 (5 μM) (*n* = 3). **e**, Treatment with SR1001 suppresses the clinical severity of EAE. Vehicle (circles, *n* = 12) or SR1001 (25 mg kg^{-1}) (triangles, *n* = 10). **f**, *Il17a*, *Il21* and *Il22* mRNA expression from spinal cords of sham-operated (IFA, incomplete Freund's adjuvant), vehicle control (V) or drug treated (SR1001) mice. (*n* = 4). All error bars denote s.e.m. *, *P* < 0.05; **, *P* < 0.01; ***, *P* < 0.001.

and *Il22* mRNA expression (Fig. 4f). Intracellular cytokine analysis of splenocytes indicated a significant reduction in IL-17 expression and reduced total $CD4^+$ T cells with no effect on $CD8^+$ T cells. Messenger RNA expression of IL-4 and IFN-γ was unaffected in both spleen and spinal cord (Supplementary Fig. 8). These data are consistent with our interpretation that SR1001 suppresses EAE through its effects on T_H17 cell function *in vivo*. Further optimization of SR1001 may yield compounds with greater activity.

Although RORα and RORγt expression and activity are essential for full T_H17 cell development, it is important to note that RORα and RORγ have roles outside of the immune system and are critical regulators of hepatic metabolism. Administration of SR1001 to C57BL/6 mice suppressed the expression of hepatic ROR target genes *Cyp7b1*, *Nr1d1* (also called *Rev-erbα*) and *Serpine1* (also called *PAI-1*), indicating that this class of compound may have metabolic effects; however, we noted no obvious toxicity in animals treated with SR1001 (Supplementary Fig. 9)[16–18].

We describe a novel, first-in-class, highly selective drug targeting the orphan nuclear receptors RORα and RORγ that effectively suppresses T_H17 cell differentiation and cytokine expression and reduces the severity of disease in an animal model of multiple sclerosis. Our data indicate that the targeting of T_H17 cells, by blocking RORα/γ function with a synthetic ligand, is a tractable approach for potential therapeutic intervention. Current treatments for T_H17-mediated autoimmune diseases, including multiple sclerosis, use agents that are general immunosuppressants and thus the side-effect profile is significant. Clearly, our data demonstrate that by targeting RORα and RORγ one can specifically inhibit T_H17 cells without affecting other T-helper cell lineages, thereby providing a more focused therapy that will not be a general immunosuppressant. Therefore, SR1001 and derivatives of this compound may represent a novel class of superior drugs to not only treat T_H17-mediated autoimmune disorders but ROR-mediated metabolic disorders as well.

METHODS SUMMARY

For the synthesis of SR1001 (*N*-(5-(*N*-(4-(1,1,1,3,3,3-hexafluoro-2-hydroxypropan-2-yl)phenyl)sulphamoyl)-4-methylthiazol-2-yl)acetamide), a solution of 2-(4-aminophenyl)-1,1,1,3,3,3-hexafluoropropan-2-ol (0.88 g, 3.4 mmol), 2-acetamido-4-methylthiazole-5-sulphonyl chloride (0.79 g, 3.1 mmol) in acetone (15 ml) and 2,6-lutidine (0.73 ml, 6.2 mmol) was warmed to 60 °C for 18 h. The reaction was judged complete by analytical HPLC (starting materials consumed).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. McGeachy, M. J. & Cua, D. J. Th17 cell differentiation: The long and winding road. *Immunity* **28,** 445–453 (2008).
2. Bettelli, E., Korn, T., Oukka, M. & Kuchroo, V. K. Induction and effector functions of T_H17 cells. *Nature* **453,** 1051–1057 (2008).
3. Littman, D. R. & Rudensky, A. Y. Th17 and regulatory T cells in mediating and restraining inflammation. *Cell* **140,** 845–858 (2010).
4. Yang, X. X. O. *et al.* T helper 17 lineage differentiation is programmed by orphan nuclear receptors RORα and RORγ. *Immunity* **28,** 29–39 (2008).
5. Ivanov, I. I. *et al.* The orphan nuclear receptor RORγt directs the differentiation program of proinflammatory IL-17+ T helper cells. *Cell* **126,** 1121–1133 (2006).
6. Ivanov, I. I., Zhou, L. & Littman, D. R. Transcriptional regulation of Th17 cell differentiation. *Semin. Immunol.* **19,** 409–417 (2007).
7. Manel, N., Unutmaz, D. & Littman, D. R. The differentiation of human T_H-17 cells requires transforming growth factor-β and induction of the nuclear receptor RORγt. *Nature Immunol.* **9,** 641–649 (2008).
8. Kumar, N. *et al.* The benzenesulfonamide T0901317 is a novel RORα/γ inverse agonist. *Mol. Pharmacol.* **77,** 228–236 (2010).
9. Zhang, F. P., Meng, G. X. & Strober, W. Interactions among the transcription factors Runx1, RORγt and Foxp3 regulate the differentiation of interleukin 17-producing T cells. *Nature Immunol.* **9,** 1297–1306 (2008).
10. Ichiyama, K. *et al.* Foxp3 inhibits RORγt-mediated IL-17A mRNA transcription through direct interaction with RORγt. *J. Biol. Chem.* **283,** 17003–17008 (2008).
11. Wang, Y. *et al.* Modulation of retinoic acid receptor-related orphan receptor α and γ activity by 7-oxygenated sterol ligands. *J. Biol. Chem.* **285,** 5013–5025 (2010).
12. Chopra, A. R. *et al.* Absence of the SRC-2 coactivator results in a glycogenopathy resembling Von Gierke's disease. *Science* **322,** 1395–1399 (2008).
13. Jin, L. H. *et al.* Structural basis for hydroxycholesterols as natural ligands of orphan nuclear receptor RORγ. *Mol. Endocrinol.* **24,** 923–929 (2010).
14. Xu, J., Wagoner, G., Douglas, J. C. & Drew, P. D. Liver X receptor agonist regulation of Th17 lymphocyte function in autoimmunity. *J. Leukoc. Biol.* **86,** 401–409 (2009).
15. Xu, J. H., Racke, M. K. & Drew, P. D. Peroxisome proliferator-activated receptor-α agonist fenofibrate regulates IL-12 family cytokine expression in the CNS: relevance to multiple sclerosis. *J. Neurochem.* **103,** 1801–1810 (2007).
16. Delerive, P., Chin, W. W. & Suen, C. S. Identification of Reverbα as a novel RORα target gene. *J. Biol. Chem.* **277,** 35013–35018 (2002).
17. Wada, T. *et al.* Identification of oxysterol 7α-hydroxylase (*Cyp7b1*) as a novel retinoid-related orphan receptor α (RORα) (NR1F1) target gene and a functional cross-talk between RORα and liver X receptor (NR1H3). *Mol. Pharmacol.* **73,** 891–899 (2008).
18. Wang, J., Yin, L. & Lazar, M. A. The orphan nuclear receptor Rev-erbα regulates circadian expression of plasminogen activator inhibitor type 1. *J. Biol. Chem.* **281,** 33842–33848 (2006).

METHODS

SR1001. For the synthesis of SR1001 (*N*-(5-(*N*-(4-(1,1,1,3,3,3-hexafluoro-2-hydroxypropan-2-yl)phenyl)sulphamoyl)-4-methylthiazol-2-yl)acetamide), a solution of 2-(4-aminophenyl)-1,1,1,3,3,3-hexafluoropropan-2-ol (0.88 g, 3.4 mmol), 2-acetamido-4-methylthiazole-5-sulphonyl chloride (0.79 g, 3.1 mmol) in acetone (15 ml) and 2,6-lutidine (0.73 ml, 6.2 mmol) was warmed to 60 °C for 18 h. The reaction was judged complete by analytical HPLC (starting materials consumed). The solvent was removed *in vacuo*, and the crude residue was diluted with EtOAc and aqueous 1 M HCl. The layers were separated, and the organic layer was washed with 1 M HCl, saturated aqueous NaHCO₃, brine, dried MgSO₄, and concentrated to give a solid. Trituration with warm Et₂O/hexanes afforded *N*-(5-(*N*-(4-(1,1,1,3,3,3-hexafluoro-2-hydroxypropan-2-yl)phenyl)sulphamoyl)-4-methylthiazol-2-yl)acetamide (1.2 g, 86% yield) as a light tan solid, >95% pure as judged by analytical HPLC. A small amount of this was further purified by reverse-phase preparative HPLC to >99% purity to give a colourless solid. ^{1}H NMR (DMSO-d_6, 400 MHz) δ 12.5 (s, 1H), 10.8 (s, 1H), 8.6 (s, 1H), 7.60 (d, 2H), 7.25 (d, 2H), 2.30 (s, 3H), 2.15 (s, 3H); ^{13}C NMR (DMSO-d_6, 100 MHz) δ 170.0, 159.6, 153.0, 139.4, 128.4, 126.8, 124.8, 121.9, 121.7, 120.3, 22.8, 16.3; ^{19}F NMR (DMSO-d_6, 376 MHz) δ −74.1; HRMS (ESI-orbitrap) calculated for $C_{15}H_{14}F_6N_3O_4S_2$, 478.0330; found, 478.0319.

Mice. All mice were maintained in a specific pathogen-free environment in accordance with institutional protocol. All procedures were reviewed and approved by either The Scripps Research Institute or the University of Arkansas for Medical Sciences Institutional Animal Care and Use Committee. C57BL/6 mice purchased from Jackson laboratories were used for all *in vitro* experiments unless otherwise noted. EAE was induced in 8-week-old male wild-type C57BL/6 mice purchased from Harlan. Male DIO mice, 22 weeks of age, were purchased from Jackson Laboratories and fed a high fat diet (60% kCal% fat) (Research Diets) for the duration of the study.

Induction and clinical evaluation of EAE. EAE was induced in C57BL/6 wild-type mice by subcutaneous injection over four sites in the flank with 200 μg per mouse MOG₃₅–₅₅ peptide (C S Bio Co.) in an emulsion with IFA supplemented with 2.25 mg ml⁻¹ *Mycobacterium tuberculosis*, strain H37Ra (Difco). Pertussis toxin (List Biological Laboratories) dissolved in PBS was injected intraperitoneally at 200 ng per mouse at the time of immunization (day 0) and 48 h later. Mice were scored daily on a scale of 0–6, as described previously[15]: 0, no clinical disease; 1, limp/flaccid tail; 2, moderate hindlimb weakness; 3, severe hindlimb weakness; 4, complete hindlimb paralysis; 5, quadriplegia or pre-moribund state; 6, death. All mice were 7–10 weeks of age when experiments were performed. The SR1001 was dissolved in DMSO at 25 mg ml⁻¹ and the mice were treated (intraperitoneally) with 25 mg kg⁻¹ SR1001 (1 μl g⁻¹ body weight of mouse) or vehicle (DMSO, 1 μl g⁻¹ body weight of mouse) twice per day. The treatment was started 2 days before immunization and continued until the end of experiment. Where indicated in the figure legends, mice were anaesthetized with halothane and transcardially perfused with PBS, and spinal cords were removed for RNA and protein isolation.

Cell lines. HEK293 cells and EL4 cells (American Type Culture Collection) were maintained in Dulbecco's modified Eagle's medium supplemented with 10% FBS and antibiotics (penicillin and streptomycin; Invitrogen). HepG2 cells were maintained in minimal essential medium supplemented with 10% FBS and antibiotics. PBMCs were obtained from Astarte Biologics and maintained in RPMI-1640 with 10% FBS and antibiotics.

T-cell differentiation *in vitro*. The conditions for the different T-helper cell subsets were: 20 μg ml⁻¹ anti-IL-4 (clone 30340, R&D Systems) and 20 μg ml⁻¹ anti-IFN-γ (clone H2, R&D Systems) for T_H0 (neutral conditions); 20 μg ml⁻¹ anti-IL-4, 20 ng ml⁻¹ IL-12 (R&D Systems) and 10 ng ml⁻¹ IFN-γ (R&D Systems) for T_H1 conditions; 20 μg ml⁻¹ anti-IFN-γ and 10 ng ml⁻¹ IL-4 (R&D systems) for T_H2 conditions; 10 μg ml⁻¹ anti-IFN-γ, 10 μg ml⁻¹ anti-IL-4 and 2 ng ml⁻¹ TGF-β (R&D Systems) for inducible T_reg conditions; 20 μg ml⁻¹ anti-IFN-γ, 20 μg ml⁻¹ anti-IL-4, 1 ng ml⁻¹ TGF-β and 10 ng ml⁻¹ IL-6 (R&D Systems) for T_H17 conditions. All cultures were stimulated with 1 μg ml⁻¹ anti-CD3 (eBiosciences) and 1 μg ml⁻¹ anti-CD28 (eBiosciences). For naive T-cell differentiation, CD4⁺CD25⁻CD62L^hiCD44^lo cells were FACS sorted on a BD FACSAriaII. Naive CD4⁺ T cells were activated with 5 μg ml⁻¹ plate-bound anti-CD3 and 1 μg ml⁻¹ anti-CD28 in the presence of 20 μg ml⁻¹ anti-IFN-γ, 20 μg ml⁻¹ anti-IL-4, 1 ng ml⁻¹ TGF-β and 10 ng ml⁻¹ IL-6, similar to splenocyte activation. Four to five days after activation, all cells were re-stimulated with 5 ng ml⁻¹ phorbol-12-myristate-13-acetate (PMA) (Sigma) and 500 ng ml⁻¹ ionomycin (Sigma) for 2 h with the addition of GolgiStop (BD Bioscience) for an additional 2 h before intracellular staining. Similar re-stimulation with PMA and ionomycin occurred for ELISA. For human PBMC analysis, cells were re-stimulated with 5 ng ml⁻¹ PMA (Sigma) and 500 ng ml⁻¹ ionomycin (Sigma) for 2 h with the addition of GolgiStop (BD Bioscience) for an additional 2 h before intracellular staining. Cells were cultured in RPMI 1640 medium (Invitrogen) with 10% FBS and antibiotics.

Reporter constructs. Gal4–RORα LBD and Gal4–RORγ LBD were gifts from Phenex Pharmaceuticals AG. The IL-17 reporter construct was purchased from ATCC and has been previously described[9].

Reporter gene assays. Twenty-four hours before transfection, HEK293 cells were plated in 96-well plates at a density of 15 × 10³ cells per well. Transfections were performed using Lipofectamine 2000 (Invitrogen). Twenty-four hours after transfection, the cells were treated with vehicle or compound. Twenty-four hours after treatment, the luciferase activity was measured using the Dual-Glo luciferase assay system (Promega). Results were analysed using GraphPad Prism software.

Radioligand binding assay. Radioligand binding assays were performed as previously described[11]. For the competition assay, various concentrations of SR1001 were incubated with receptor in the presence of 3 nM [³H]-25-hydroxycholesterol. Results were analysed using GraphPad Prism software and the K_i was determined using the Cheng–Prusoff equation.

Alpha screen. The Alpha screen assays were performed as previously described[8]. Assays were performed in triplicate in white opaque 384-well plates (Greiner Bio-One) under green light conditions (<100 lx) at room temperature. The final assay volume was 20 μl. All dilutions were made in assay buffer (100 mM NaCl, 25 mM HEPES, 0.1% BSA, pH 7.4). The final DMSO concentration was 0.25%. A mix of 12 μl of GST–RORγ LBD (10 nM), beads (12.5 μg ml⁻¹ of each donor and acceptor), and 4 μl of increasing concentrations (210 nM to 50 μM) of compound SR1001 were added to the wells, the plates were sealed and incubated for 1 h. After this pre-incubation step, 4 μl of biotin-TRAP220-2 peptide (50 nM) was added, the plates were sealed and further incubated for 2 h. The plates were read on PerkinElmer Envision 2104 and data analysed using GraphPad Prism software.

RNA-mediated interference. EL4 cells were first electroporated with 100 nM total siRNA with the GenePulserXcell Electroporator using siRNA against mouse RORα and RORγ (Dharmacon RNA Technologies) followed by reverse transfection using 50 nM siRNA according to the instructions for Dharma-FECT 1 transfection reagent and seeded onto a 12-well plate. Twenty-four hours after transfection, cells were treated with vehicle (DMSO) or SR1001 (10 μM) for 24 h. Cells were harvested and total RNA was isolated. Quantitative reverse transcriptase PCR was performed to analyse mRNA levels of mouse *Rora*, *Rorc*, *Gapdh* and *Il17a* using SYBR green technology. Primers sequences to mouse *Rora*, *Rorc*, *Il17a* and *Gapdh* and have previously been described[5,6,19]. HepG2 cells were treated similarly to EL4 cells with the following exceptions: HepG2 cells were transfected with siRNA against human *RORA* and *RORC* (Dharmacon RNA Technologies) at 50 nM according to the instructions for Dharma-FECT 1 transfection reagent. Quantitative reverse transcriptase PCR was performed to analyse mRNA levels of human *RORA*, *RORC*, cyclophilin and *G6PC* using SYBR green technology. The primer sequences have previously been described[8].

Quantitative RT–PCR. Splenocyte total RNA was extracted using a RNeasy Plus Micro Kit (Qiagen) and reverse transcribed using the iScript cDNA biosynthesis kit (Bio-Rad). Total RNA from spinal cord from MOG-immunized mice and livers from DIO mice was isolated using TRIzol reagent (Invitrogen) followed by DNase1 treatment (Invitrogen) and reverse transcribed using the iScript cDNA biosynthesis kit. Quantitative RT–PCR was performed with a 7900HT Fast Real Time PCR System (Applied Biosystems) using SYBR green (Roche) as previously described[20]. Primer sequences for mouse *Il17a*, *Rorc*, *Rora*, *Il17f*, *Il21*, *Il22*, T-bet (also called *Tbx21*), *Gata3* and *Foxp3* have previously been described[5,19,21,22]. The level of mRNA expression was normalized to that of *Gapdh* mRNA expression. The following primer sequences were used: *Cyp7b1* forward 5′-GGGAAG AAGCTGAAGACTTACG-3′, reverse 5′-CCCTATAGGCTTCCTGTCGAT-3′; *Nr1d1* forward 5′-ACCTTTGAGGTGCTGATGGT-3′, reverse 5′-CTCGCTG AAGTCAAACATGG-3′; *Serpine1* forward 5′-CTCGCTGAAGTCAAACAT GG-3′, reverse 5′-TTTTGCAGTGCCTGTGCTAC-3′.

Western blot analysis. HepG2 and EL4 cells were washed once with phosphate-buffered saline and then incubated for 10 min at 4 °C in 100 μl of TNT lysis buffer (50 mM Tris-Cl, pH 7.5, 150 mM NaCl and 1% Triton X-100) and a complete miniprotease inhibitor mixture (Roche Applied Science). Samples were then harvested into 1.5-ml microcentrifuge tubes, vortexed for 30 s, and then centrifuged (10 min). Protein levels in the supernatants were determined using a Coomassie protein assay kit (Bio-Rad), and 10 μg of protein from each sample was separated by SDS–PAGE (BioRad, 10%) and then transferred to a polyvinylidene difluoride membrane (Millipore) and immunoblotted with primary antibodies: mouse RORα (BioLegend), mouse RORγ (BioLegend), human RORα (Perseus Proteomics), human RORγ (IMGENEX), or α-tubulin (Sigma) and horseradish peroxidase-conjugated secondary antibodies (Jackson Immunoresearch). Detection of the bound antibody by enhanced chemiluminescence was performed according to the manufacturer's instructions (Santa Cruz).

Hydrogen/deuterium exchange mass spectrometry. Differential, solution phase HDX experiments were performed with a LEAP Technologies Twin HTS PAL liquid handling robot interfaced with an Orbitrap mass spectrometer (Exactive,

ThermoFisher Scientific)[23]. Each exchange reaction was initiated by incubating 4 µl of 10 µM protein complex (with or without SR1001 and SRC2 RID) with 20 µl of D_2O protein buffer for a predetermined time (1 s, 30 s, 60 s, 900 s and 3,600 s) at 25 °C. The exchange reaction was quenched by mixing with 50 µl of 3 M urea, 0.1% TFA at 1 °C. The mixture was passed across at an in-house packed pepsin column (2 mm × 2 cm) at 50 µl min^{-1} and digested peptides were captured onto a 2 mm × 1 cm C_8 trap column (Agilent) and desalted (total time for digestion and desalting was 2.5 min). Peptides were then separated across a 2.1 mm × 5 cm C_{18} column (1.9µ Hypersil Gold, Thermo Scientific) with a linear gradient of 4–40% CH_3CN, 0.3% formic acid, over 5 min. Protein digestion and peptide separation were performed within a thermal chamber (Mécour) held at 2 °C to reduce D/H back exchange. Mass spectrometric analyses were carried out with capillary temperature at 225 °C and data were acquired with a measured resolving power of 65,000 at m/z 400. Three replicates were performed for each on-exchange time point.

Peptide identification and HDX data processing. MS/MS experiments were performed with a linear ion trap mass spectrometer (LTQ, ThermoFisher). Product ion spectra were acquired in a data-dependent mode and the five most abundant ions were selected for the product ion analysis. The MS/MS *.raw data files were converted to *.mgf files and then submitted to Mascot (Matrix Science) for peptide identification. Peptides included in the peptide set used for HDX had a MASCOT score of 20 or greater. The MS/MS MASCOT search was also performed against a decoy (reverse) sequence and ambiguous identifications were ruled out. The MS/MS spectra of all of the peptide ions from the MASCOT search were further manually inspected and only those verifiable are used in the coverage. The intensity weighted average m/z value (centroid) of each peptide isotopic envelopes were calculated with a new version of our in-house developed software; HD Desktop[24]. The deuterium level was calculated as described previously[25]. Deuterium level (%) = ((m(P)−m(N))/(m(F)-m(N)))×100%, where m(P), m(N) and m(F) are the centroid value of partly deuterated peptide, non-deuterated peptide and fully deuterated peptide, respectively. The corrections for back-exchange were made based on an estimated 70% deuterium recovery and accounting for the known 80% deuterium content of the on-exchange buffer.

ChIP-ReChIP. EL4 cells were treated with plate-bound anti-CD3 (5 µg ml^{-1}) and soluble anti-CD28 (1 µg ml^{-1}) for 24 h and then treated with vehicle (DMSO) or SR1001 (10 µM) for another 24 h. Re-ChIP assays were performed by using the kit from Active Motif Inc. (Carlsbad). Anti-RORα (BioLegend) or anti-RORγ (BioLegend) antibody was used to do the first immunoprecipitation for all of the samples. The second immunoprecipitation was performed by using anti-rabbit IgG (Jackson Immunoresearch) for RORα, anti-hamster IgG (Jackson Immunoresearch) for RORγ, anti-RNA Pol II (Millipore), anti-SRC2 (Bethyl Laboratories), or anti-NCoR (Santa Cruz). The IL-17 primers used in PCR have been previously described[4].

Flow cytometry and antibodies. For the analysis of all T-cell subsets, single-cell suspensions prepared from spleen were stained with fluorophore-conjugated monoclonal antibodies: FITC anti-CD4 (GK1.5, eBioscience) and Alexafluor 647 anti-Foxp3 (FJK-16 s, eBioscience) along with the Foxp3 staining buffer set (eBioscience). For intracellular cytokine staining, phycoerythrin-conjugated anti-mouse IL-17A (eBio17B7, eBioscience), phycoerythrin-conjugated anti-mouse IL-4 (11B11, eBioscience), and Alexafluor 647 anti-mouse IFN-γ (XMG1.2, eBioscience), in conjunction with the BD Bioscience Intracellular Cytokine staining kit, were used. For the intracellular cytokine analysis of human PBMCs, FITC anti-CD4 (RPA-T4, eBioscience) and phycoerythrin-conjugated anti-human IL-17A (eBio64DEC17, eBioscience) were used. For the sorting of naive CD4$^+$ T cells, CD4$^+$CD25$^-$CD62L^{hi}CD44^{lo} cells were gated on using the following fluorophore-conjugated antibodies: FITC anti-CD4 (GK1.5, eBioscience), phycoerythrin-conjugated anti-CD25 (PC61.5, eBioscience), APC-conjugated anti-CD62L (MEL-14, eBioscience), and Alexa Fluor 700 anti-CD44 (IM7, eBioscience). Flow cytometric analysis was performed on a BD LSRII (BD Biosciences) instrument and analysed using FlowJo software (TreeStar).

ELISA. The concentration of IL-17 in the culture supernatant was determined by an ELISA kit according to the manufacturers protocol (R&D Systems).

Statistical analysis. All data are expressed as the mean ± s.e.m. ($n = 3$ or more). Statistical analysis was performed using an unpaired Student's t-test.

19. Okamoto, K. *et al.* IκBζ regulates T_H17 development by cooperating with ROR nuclear receptors. *Nature* **464,** 1381–1385 (2010).
20. Raghuram, S. *et al.* Identification of heme as the ligand for the orphan nuclear receptors REV-ERBα and REV-ERBβ. *Nature Struct. Mol. Biol.* **14,** 1207–1213 (2007).
21. Lighvani, A. A. *et al.* T-bet is rapidly induced by interferon-γ in lymphoid and myeloid cells. *Proc. Natl Acad. Sci. USA* **98,** 15137–15142 (2001).
22. Tamauchi, H. *et al.* Evidence of GATA-3-dependent T_h2 commitment during the *in vivo* immune response. *Int. Immunol.* **16,** 179–187 (2004).
23. Chalmers, M. J. *et al.* Probing protein ligand interactions by automated hydrogen/deuterium exchange mass spectrometry. *Anal. Chem.* **78,** 1005–1014 (2006).
24. Pascal, B. D., Chalmers, M. J., Busby, S. A. & Griffin, P. R. H. D. Desktop: An integrated platform for the analysis and visualization of H/D exchange data. *J. Am. Soc. Mass Spectrom.* **20,** 601–610 (2009).
25. Zhang, Z. Q. & Smith, D. L. Determination of amide hydrogen-exchange by mass-spectrometry—a new tool for protein-structure elucidation. *Protein Sci.* **2,** 522–531 (1993).

LETTER

Differential microRNA regulation of *HLA−C* expression and its association with HIV control

Smita Kulkarni[1,2]*, Ram Savan[3]*, Ying Qi[1,2], Xiaojiang Gao[1,2], Yuko Yuki[1,2], Sara E. Bass[1], Maureen P. Martin[1,2], Peter Hunt[4], Steven G. Deeks[4], Amalio Telenti[5], Florencia Pereyra[2], David Goldstein[6], Steven Wolinsky[7], Bruce Walker[2], Howard A. Young[3] & Mary Carrington[1,2]

The *HLA-C* locus is distinct relative to the other classical *HLA* class I loci in that it has relatively limited polymorphism[1], lower expression on the cell surface[2,3], and more extensive ligand–receptor interactions with killer-cell immunoglobulin-like receptors[4]. A single nucleotide polymorphism (SNP) 35 kb upstream of *HLA-C* (rs9264942; termed −35) associates with control of HIV[5–7], and with levels of *HLA-C* messenger RNA transcripts[8] and cell-surface expression[7], but the mechanism underlying its varied expression is unknown. We proposed that the −35 SNP is not the causal variant for differential HLA-C expression, but rather is marking another polymorphism that directly affects levels of HLA-C[7]. Here we show that variation within the 3′ untranslated region (UTR) of *HLA-C* regulates binding of the microRNA hsa-miR-148 to its target site, resulting in relatively low surface expression of alleles that bind this microRNA and high expression of *HLA-C* alleles that escape post-transcriptional regulation. The 3′ UTR variant associates strongly with control of HIV, potentially adding to the effects of genetic variation encoding the peptide-binding region of the *HLA* class I loci. Variation in HLA-C expression adds another layer of diversity to this highly polymorphic locus that must be considered when deciphering the function of these molecules in health and disease.

MicroRNAs (miRNAs) are a class of non-protein-coding RNAs that are estimated to regulate 30% of all genes in animals[9] by binding to specific sites in the 3′ UTR, resulting in post-transcriptional repression, cleavage or destabilization[10–12]. The 3′ UTR of the *HLA-C* gene is predicted to be a target for 26 distinct human miRNAs using three miRNA-target-prediction programs (Supplementary Fig. 1), of which three (miR-148a and miR-148b, which bind the same target site, and miR-657) were shown to have the greatest likelihood of binding. We sequenced the 3′ UTRs of the common *HLA-C* alleles (Supplementary Fig. 2) and show that the two binding sites of these three miRNAs are polymorphic (Supplementary Fig. 3a). The binding site for miR-148a/miR-148b contains a single base pair insertion/deletion at position 263 downstream of the *HLA-C* stop codon (rs67384697G representing the insertion (263ins) and rs67384697− representing the deletion (263del)) along with other precisely linked variants (259C/T, 261T/C, 266C/T). These variants are likely to impose a restriction in miR-148a/miR-148b binding, as prediction algorithms indicate that the binding of these miRNAs to the alleles marked by 263ins (for example, *Cw*0702*, a low-expression allotype) is more stable than to alleles with 263del (for example, *Cw*0602*, a high-expression allotype) (Supplementary Fig. 3b). Similarly, alleles with 307C within the miR-657 target site are predicted to be better targets of miR-657 than those with 307T (Supplementary Fig. 4). Thus, variation in the 3′ UTR of *HLA-C* may influence the interaction between these miRNAs and their

putative binding sites in an allele-specific manner, potentially leading to differential levels of HLA-C allotype expression.

To test directly whether the variation in the *HLA-C* 3′ UTR affects levels of protein expression, the full-length 3′ UTRs containing intact miR-148a/miR-148b- and miR-657-binding sites (that is, 263ins and 307C, respectively; *Cw*0702*, *Cw*0303*, *Cw*0401*, *Cw*0701*) and disrupted binding sites (that is, 263del and 307T, respectively; *Cw*0602*, *Cw*0802*, *Cw*1203*, *Cw*1502*) were each cloned downstream of the luciferase gene in a pGL3 reporter construct (Fig. 1a). The constructs were then transfected into HLA class I negative B721.221 cells, and the
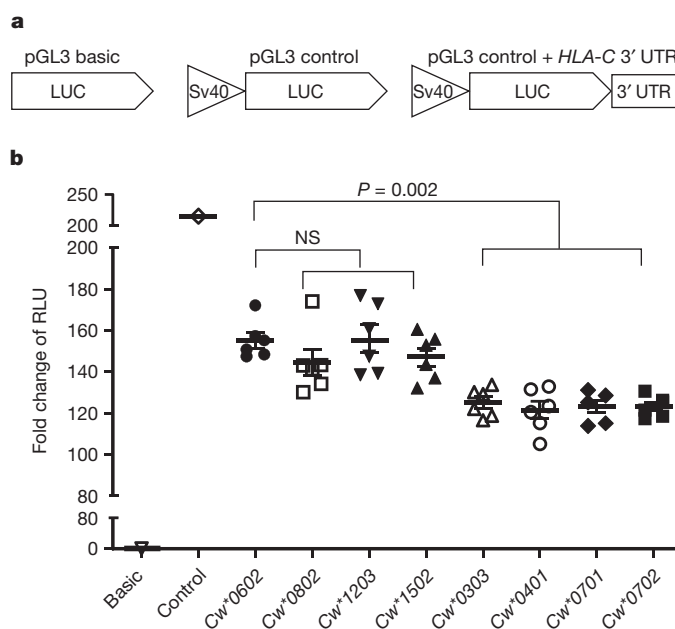

Figure 1 | Variation in the *HLA-C* 3′ UTR differentially affects the expression of a reporter gene. a, b, Full-length 3′ UTRs of various *HLA-C* alleles cloned into luciferase (LUC) reporter constructs were transfected into B721.221 cell lines and the stability of the mRNA was estimated by dual luciferase reporter assays. The normalized luciferase activity is presented as fold change of relative light units (RLU). The data represent six replicates in each experimental group, the mean ± standard error (s.e.) are depicted as horizontal and vertical bars for each group, respectively, and one of three comparable experiments performed is shown. Non-parametric Wilcoxon–Mann–Whitney tests were used for statistical comparisons and two-tailed *P* values are indicated. NS, not significant. **a,** Schematic representations of the luciferase reporter constructs used in this study. Sv40, Simian virus 40. **b,** Fold change in luciferase activity of 3′ UTRs of *HLA-C* alleles as compared to that of *Cw*0602*.

[1]Cancer and Inflammation Program, Laboratory of Experimental Immunology, SAIC-Frederick, Inc., NCI-Frederick, Frederick, Maryland 21702, USA. [2]Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology and Harvard University, Boston, Massachusetts 02114, USA. [3]Cancer and Inflammation Program, Laboratory of Experimental Immunology, NCI-Frederick, Frederick, Maryland, 21702 USA. [4]San Francisco General Hospital AIDS division, University of California, San Francisco, California 94110, USA. [5]Institute of Microbiology, University of Lausanne, Lausanne 1011, Switzerland. [6]Institute for Genome Sciences & Policy, Center for Human Genome Variation, Duke University, Durham, North Carolina 27708, USA. [7]Division of Infectious Diseases, The Feinberg School of Medicine, Northwestern University, Chicago, Illinois 60611, USA.
*These authors contributed equally to this work.

level of luciferase activity was measured (fold increase of relative light units). Although the *Cw*0602* 3′ UTR repressed luciferase activity as compared to the control containing no 3′ UTR, the constructs containing intact miRNA-binding sites (that is, 263ins and 307C; *Cw*0702*, *Cw*0303*, *Cw*0401*, *Cw*0701*) produced significantly lower luciferase activity relative to the construct containing the 3′ UTR of *Cw*0602*, which contains 263del and 307T (Fig. 1b). However, 3′ UTRs from other alleles with the 263del and 307T variants (*Cw*0802*, *Cw*1203*, *Cw*1502*) did not show significant variation in luciferase activity as compared to *Cw*0602* (Fig. 1b). Psicheck2 reporter constructs containing 3′ UTRs of *Cw*0602* also produced significantly higher luciferase activity as compared to those with *Cw*0702* 3′ UTR (Supplementary Fig. 5a), indicating that this effect was reproducible in a distinct reporter construct. Further, pGL3 constructs containing 3′ UTRs of *Cw*0602* and *Cw*0702* in three additional cell lines showed the same pattern as that seen in B721.221 cells, indicating a consistent difference of these 3′ UTRs in the regulation of HLA-C expression that is independent of cell type (Supplementary Fig. 5b–e). Thus, *HLA-C* 3′ UTR alleles characterized by variation at positions 263 and 307 within miRNA-binding regions differentially regulate gene expression.

The expression of endogenous mature miR-148b and miR-152, another miR-148 miRNA family member, was very low as compared to miR-148a, and miR-657 was undetectable in *HLA-C* homozygous B lymphoblastoid cell lines (BLCLs) and B721.221 cells (Supplementary Fig. 6). These data point to the involvement of miR-148a rather than miR-148b, miR-152, or miR-657 in regulation of HLA-C expression. Additionally, disruption of the miR-657-binding site by site-directed mutagenesis had no effect on luciferase activity (Supplementary Fig. 7a, b), indicating that miR-657 does not affect HLA-C expression.

To test whether variants in the miR-148a-binding site account for the differential gene expression patterns, we swapped positions 256–266 of the 3′ UTR of *Cw*0602* to match those of *Cw*0702* and vice versa, thereby providing an intact miR-148a-binding site to the 3′ UTR

of *Cw*0602* (*06mut*) and disrupting the binding site for miR-148a in the *Cw*0702* 3′ UTR (*07mut*), but leaving the remainder of the 3′ UTR sequences intact (Fig. 2a). The luciferase activity of *06mut* was significantly lower than that of *07mut* (Fig. 2b), indicating that the polymorphisms between positions 256–266 in the miR-148a-binding region account for the difference in luciferase expression between constructs containing the 3′ UTRs of *Cw*0602* versus *Cw*0702*. Two other polymorphic sites, A256C and A267G, in the miR-148a-binding site (Supplementary Fig. 3a) distinguish different sets of alleles as compared to 263del/ins, but these two variants had no effect on miRNA-mediated suppression (Supplementary Fig. 7c, d).

Further validation of the differential regulation of *HLA-C* alleles by miR-148a was achieved by co-transfection of B221.227 cell lines with either a mimic or an inhibitor of miR-148a along with a luciferase reporter construct that contained the 3′ UTR with 263ins (*Cw*0702* or *06mut*) or with 263del (*Cw*0602* or *07mut*). The normalized luciferase activity in cells transfected with the constructs containing the 263del allele (*Cw*0602* and *07mut* 3′ UTR) was not significantly altered by co-transfection with either the mimic or the inhibitor (Fig. 2c, d). However, the mimic of miR-148a further repressed luciferase activity in cells transfected with the 263ins allele (*Cw*0702* and *06mut* 3′ UTR), whereas co-transfection with inhibitor rescued the suppression significantly (Fig. 2c, d). These data provide further support for allele-specific miR-148a targeting of the *HLA-C* 3′ UTR.

BLCLs from individuals homozygous for either *Cw*0602* (BLCL-Cw0602Hom) or *Cw*0702* (BLCL-Cw0702Hom) were used to determine whether the variation in the miR-148a-binding site affected endogenous HLA-C expression on the cell surface. As described previously, overall HLA-C expression on a *Cw*0602* homozygous cell line was higher than that on a *Cw*0702* homozygous cell line. As expected, transfection with mimics or inhibitors of miR-148a (Fig. 3a) and miR-148b (Fig. 3b) had no significant effect on cell-surface expression of Cw*0602, an allele containing 263del in the 3′ UTR that disrupts
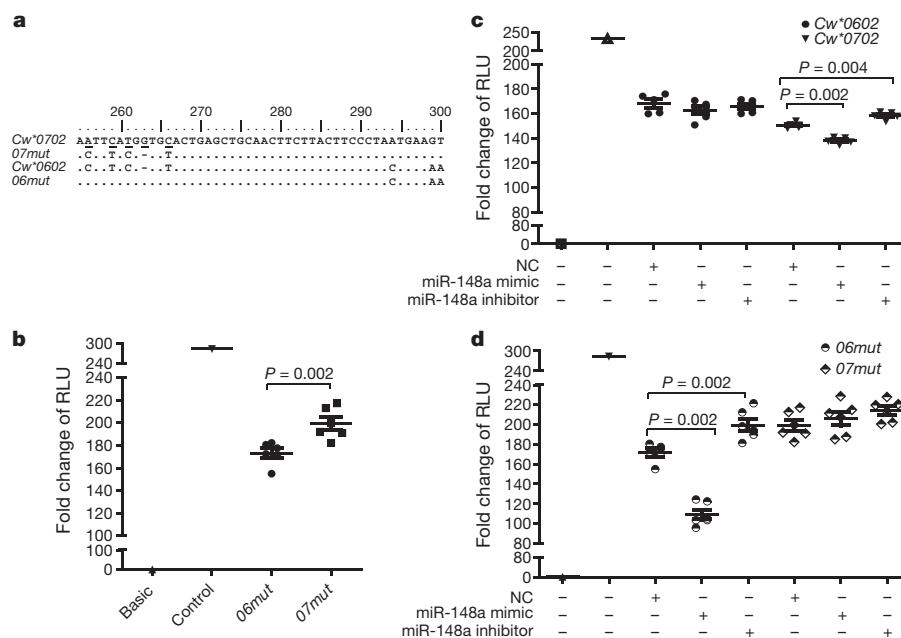

Figure 2 | Disruption of miR-148a target site rescues suppression. a, Partial sequence of mutated 3′ UTRs of *Cw*0602* and *Cw*0702* (*06mut* and *07mut*, respectively) are aligned to 3′ UTR sequences of native *Cw*0602* and *Cw*0702*. Identical nucleotides are shown as dots, altered nucleotides are underlined, and deletions are indicated by dashes for optimal alignment. **b**, Fold change in luciferase activity of the modified 3′ UTR (*06mut* and *07mut*). **c**, Fold change in luciferase activity of reporters containing wild-type *Cw*0602* or *Cw*0702* 3′ UTR sequences upon introduction of miR-148a mimic and inhibitor. **d**, Fold change in luciferase activity of reporters containing *06mut* and *07mut* 3′ UTR

sequences upon introduction of miR-148a mimic and inhibitor. Presence (+) or absence (−) of each variable, including a negative control (NC) miRNA, a mimic of miR-148a, or an inhibitor of miR-148a is shown. The data represent six replicates in each experimental group, the mean ± s.e. are depicted as horizontal and vertical bars for each group, respectively, and one of three comparable experiments performed is shown. Non-parametric Wilcoxon–Mann–Whitney tests were used for statistical comparisons and two-tailed P values are indicated.
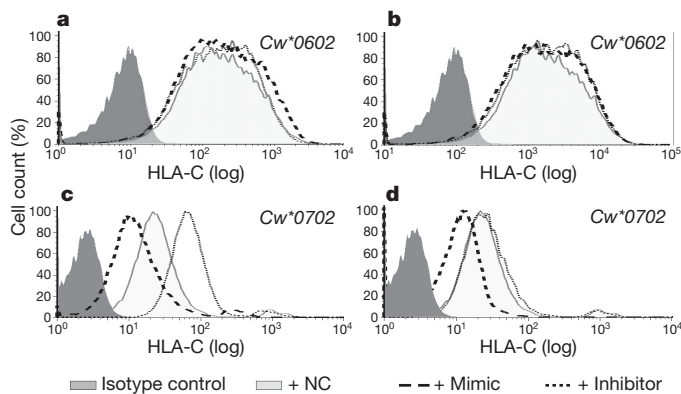
Figure 3 | miR-148a affects cell-surface expression of HLA-C.
a–d, Histograms of HLA-C cell-surface expression on HLA-C homozygous BLCLs using flow cytometry are illustrated. In each plot, a negative control (NC) miRNA that does not bind to the 3′ UTR of *HLA-C* was included. **a, b**, *HLA-Cw*0602* homozygous cells (BLCL-Cw*0602Hom) transfected with either a mimic or an inhibitor of miR-148a (**a**) or miR-148b (**b**). **c, d**, *HLA-Cw*0702* homozygous cells (BLCL-Cw*0702Hom) transfected with either a mimic or an inhibitor of miR-148a (**c**) or miR-148b (**d**).

miR-148a/miR-148b binding. However, transfection of either miR-148a or miR-148b mimic resulted in decreased expression of Cw*0702 relative to cells transfected with a negative control (Fig. 3c and d, respectively), indicating that increased levels of either of these miRNAs can further downregulate HLA-C expression of alleles that contain an intact binding site for miR-148a/miR-148b (263ins), such as Cw*0702. The inhibitor of miR-148a significantly increased the level of endogenous Cw*0702 expression, but the inhibitor of miR-148b had no effect on expression of Cw*0702 (Fig. 3c and d, respectively), confirming the very low levels of miR-148b endogenous expression (Supplementary Fig. 6).

The differential miR-148a regulation of expression across *HLA-C* alleles was precisely reflected in experiments involving: (1) additional *HLA-C* homozygous BLCLs (Supplementary Fig. 8a, b); (2) use of another form of miR-148a inhibitor (peptide nucleic acid inhibitor; Supplementary Fig. 8c); and (3) analysis of total cellular HLA-C protein as determined by western blot (Supplementary Figs 9, 10 and Supplementary Note 1). We conclude that miR-148a regulates the expression of HLA-C in an allele-specific manner that is dependent on variation in the miR-148a-binding site of the *HLA-C* 3′ UTR.

The miR-148a-binding site of the *HLA-C* 3′ UTR is in strong linkage disequilibrium with the −35 SNP that was shown to associate with control of HIV and HLA-C expression levels[5–7] ($D' = 0.75$, $r^2 = 0.74$; $P < 0.0001$, $N = 1,760$). Although there is no explanation for a direct causal effect of −35 variation on HLA-C expression[7], the interaction between miR-148a and its polymorphic binding site in the 3′ UTR of *HLA-C* presents a clear rationale for variable levels of HLA-C expression. We determined the frequencies of the 263del/ins genotypes in a cohort of 2,527 HIV-infected European American individuals. Subjects with mean plasma viral loads of <2,000 copies of viral RNA per ml of plasma (controllers) were enriched for 263del, whereas those with viral loads of >10,000 copies of viral RNA per ml of plasma (noncontrollers) had a significantly higher frequency of 263ins (Supplementary Table 1). Because of the strong linkage disequilibrium across the *HLA-C* and *HLA-B* genes (Supplementary Tables 2, 3), we determined whether the 3′ UTR variant has an effect on HIV control that is independent of individual *HLA-A*, *B*, or *C* alleles. A logistic regression approach with stepwise selection of the *HLA-C* 3′ UTR 263 variant along with all *HLA-A*, *B* and *C* alleles that have ≥1% frequency in our cohort (63 alleles) was used. In this analysis, the 263del/del versus 263ins/ins comparison remains significant along with 5 of the 63 *HLA* alleles (*B*5701*, *B*5703*, *B*2705*, *B*5801* and *Cw*1402*; Table 1; for frequencies, see Supplementary Table 1).

Table 1 | Effect of *HLA-C* 3′ UTR 263 on mean viral load

Significant independent variables	P value	OR	95% CI
del/del versus ins/ins	2×10^{-14}	0.33	0.25–0.43
B*2705 versus others	3×10^{-6}	0.34	0.22–0.54
B*5701 versus others	1×10^{-12}	0.21	0.14–0.32
B*5703 versus others	3×10^{-5}	0.01	0.002–0.10
B*5801 versus others	9×10^{-4}	0.27	0.12–0.59
Cw*1402 versus others	1×10^{-4}	0.26	0.12–0.52

CI, confidence interval; OR, odds ratio. $N = 2,527$. A logistic regression analysis with stepwise selection using the *HLA-C* 3′ UTR 263 del/del versus ins/ins comparison and 63 *HLA-A*, *B* and *C* alleles with ≥1% frequency as independent variables in the model was performed using PROC LOGISTIC (SAS 9.1 version, SAS Institute). Significance level for selecting variables shown in the table was $P < 0.05$.

Although these data indicate that the 3′ UTR del/ins variant has an independent effect on HIV control (see Supplementary Note 2 for potential mechanisms that could explain the association), we still cannot completely rule out the possibility that the strong linkage disequilibrium in the region is confounding the results[13] (Supplementary Table 4 and Supplementary Note 3).

The extensive number of disease associations with *HLA* class I and II genes has largely been ascribed to the polymorphic peptide-binding amino acid positions of these molecules. Some reports have speculated that gene expression[14] and/or splicing patterns of the *HLA* genes[15,16] may have a role, but convincing data are missing. Of interest, the *HLA-G* 3′ UTR was shown to encode a polymorphic target site for miR-148a/b[17,18]. Levels of HLA-G have been suggested to alter risk of asthma in children of mothers with asthma[18], although the specificity of assays reporting HLA-G expression beyond implanting placental cells has been questioned[19]. Recently, a variant 35 kb upstream of the HLA-C gene was shown to associate with differential *HLA-C* mRNA expression, cell-surface expression and outcome after HIV infection[5–8]. We have now established a very convincing case that this −35 SNP is marking a functional insertion/deletion variant in the 3′ UTR of *HLA-C* that directly determines expression of the various HLA-C allotypes differentially through miR-148a recognition. These data indicate another tier of diversity to the polymorphic *HLA-C* locus beyond that encoding variation in the peptide-binding region of the gene. We suggest that disease-associated haplotypes may exert their effects through multiple mechanisms, including the type of peptides they bind and their level of expression, and that it is the combination of these that then determines the overall susceptibility status of the haplotype.

Expression levels of different HLA-C allotypes occur as a continuous gradient rather than the bimodal expression pattern that would be expected if miR-148a regulation were the sole mechanism involved. Thus, additional *cis*-acting factors may fine-tune HLA-C expression in an allotype-specific manner. *Trans*-acting factors unlinked to the *HLA-C* locus may also affect expression levels in a manner that is independent of HLA-C allotype, leading to some degree of variation in expression levels of a given HLA-C allotype. Although the system regulating *HLA-C* expression is multifactorial, the significant involvement of miRNA in this process provides new approaches for manipulation of the immune system in the treatment of human disease.

METHODS SUMMARY

DNA from 2,527 HIV[+] patients of European descent was used to determine the effect of the *HLA-C* 3′ UTR variation on control of HIV viral load. Viral load measurements were obtained from participants of the Multicentre AIDS cohort study (MACS)[20], Swiss HIV Cohort (http://www.shcs.ch), the SCOPE cohort[21] and the International HIV Controllers Study Cohort (http://www.hivcontrollers. org). Individuals were grouped into those who maintain mean viral load < 2,000 (controllers) and those who have mean viral load > 10,000 (noncontrollers).

Complete *HLA-C* 3′ UTR fragments were amplified, inserted into the XbaI site downstream of the luciferase gene in a pGL3-control vector (Promega), and transfected into B721.221, BLCL and Jurkat cells using AMAXA nucleofector (Lonza) and into 293T cells using Fugene6 (Roche). Luciferase activity was measured using the Dual Luciferase Reporter Assay System (Promega) and presented as fold change of relative light units[22]. For studies of miR-148a/miR-148b mimics and inhibitors (Dharmacon), 20 pmol per well of oligonucleotide mimics or inhibitors of miR-148a and miR-148b were transfected into the cells. Surface expression of

HLA-C on BLCLs was analysed using staining with DT9 antibody (provided by V. Braud)[23].

Total RNA was extracted using the Total RNA Purification Kit (Norgen). Relative quantification of miR-148a and miR-148b was performed using a Taqman real-time PCR assay (Applied Biosystems) and RNU48 served as an endogenous RNA control.

SAS9.1 (SAS Institute) was used for data management and statistical analyses.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Zemmour, J. & Parham, P. Distinctive polymorphism at the HLA-C locus: implications for the expression of HLA-C. *J. Exp. Med.* **176,** 937–950 (1992).
2. McCutcheon, J. A., Gumperz, J., Smith, K. D., Lutz, C. T. & Parham, P. Low HLA-C expression at cell surfaces correlates with increased turnover of heavy chain mRNA. *J. Exp. Med.* **181,** 2085–2095 (1995).
3. Snary, D., Barnstable, C. J., Bodmer, W. F. & Crumpton, M. J. Molecular structure of human histocompatibility antigens: the HLA-C series. *Eur. J. Immunol.* **7,** 580–585 (1977).
4. Bashirova, A. A., Martin, M. P., McVicar, D. W. & Carrington, M. The killer immunoglobulin-like receptor gene cluster: tuning the genome for defense. *Annu. Rev. Genomics Hum. Genet.* **7,** 277–300 (2006).
5. Fellay, J. *et al.* A whole-genome association study of major determinants for host control of HIV-1. *Science* **317,** 944–947 (2007).
6. International HIV Controllers Study. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* **310,** 1551–1557 (2010).
7. Thomas, R. *et al.* HLA-C cell surface expression and control of HIV/AIDS correlate with a variant upstream of *HLA-C. Nature Genet.* **41,** 1290–1294 (2009).
8. Stranger, B. E. *et al.* Genome-wide associations of gene expression variation in humans. *PLoS Genet.* **1,** e78 (2005).
9. Lewis, B. P., Burge, C. B. & Bartel, D. P. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120,** 15–20 (2005).
10. Giraldez, A. J. *et al.* Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312,** 75–79 (2006).
11. Lim, L. P. *et al.* Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433,** 769–773 (2005).
12. Wu, L., Fan, J. & Belasco, J. G. MicroRNAs direct rapid deadenylation of mRNA. *Proc. Natl Acad. Sci. USA* **103,** 4034–4039 (2006).
13. Corrah, T. W. *et al.* A reappraisal of the relationship between the HIV-1-protective single nucleotide polymorphism 35 kb upstream of the *HLA-C* gene and surface HLA-C expression. *J. Virol.* **85,** 3367–3374 (2011).
14. Schaefer, M. R. *et al.* A novel trafficking signal within the HLA-C cytoplasmic tail allows regulated expression upon differentiation of macrophages. *J. Immunol.* **180,** 7804–7817 (2008).
15. Kralovicova, J. & Vorechovsky, I. Position-dependent repression and promotion of DQB1 intron 3 splicing by GGGG motifs. *J. Immunol.* **176,** 2381–2388 (2006).
16. Krangel, M. S. Secretion of HLA-A and -B antigens via an alternative RNA splicing pathway. *J. Exp. Med.* **163,** 1173–1190 (1986).
17. Castelli, E. C. *et al. In silico* analysis of microRNAs targeting the *HLA-G* 3′ untranslated region alleles and haplotypes. *Hum. Immunol.* **70,** 1020–1025 (2009).
18. Tan, Z. *et al.* Allele-specific targeting of microRNAs to *HLA-G* and risk of asthma. *Am. J. Hum. Genet.* **81,** 829–834 (2007).
19. Apps, R., Gardner, L. & Moffett, A. A critical look at HLA-G. *Trends Immunol.* **29,** 313–321 (2008).
20. Phair, J. *et al.* Acquired immune deficiency syndrome occurring within 5 years of infection with human immunodeficiency virus type-1: the Multicenter AIDS Cohort Study. *J. Acquir. Immune Defic. Syndr.* **5,** 490–496 (1992).
21. Emu, B. *et al.* Phenotypic, functional, and kinetic parameters associated with apparent T-cell control of human immunodeficiency virus replication in individuals with and without antiretroviral treatment. *J. Virol.* **79,** 14169–14178 (2005).
22. Li, H., Wright, P. W. & Anderson, S. K. Identification and analysis of novel transcripts and promoters in the human killer cell immunoglobulin-like receptor (KIR) genes. *Methods Mol. Biol.* **612,** 377–391 (2010).
23. Braud, V. M., Allan, D. S., Wilson, D. & McMichael, A. J. TAP- and tapasin-dependent HLA-E surface expression correlates with the binding of an MHC class I leader peptide. *Curr. Biol.* **8,** 1–10 (1998).

Author Contributions S.K. and R.S. performed and evaluated the miRNA experiments. S.K., R.S., H.A.Y. and M.C. designed the study. M.C. directed the study. S.K., R.S. and M.C. wrote the manuscript. X.G., Y.Y., S.B. and M.M. genotyped *HLA*. Statistical analysis was performed by Y.Q. The clinical samples and data were contributed to by P.H., S.G.D., D.D., A.T., D.G., S.W., F.P. and B.W. Intellectual input was provided by all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.C. (carringm@mail.nih.gov)

METHODS

Subjects. DNA from 2,527 HIV$^+$ patients of European descent was used to determine the effect of the *HLA-C* 3′ UTR variation on control of HIV viral load. DNA from 1,760 individuals of European descent was used to determine the patterns of linkage disequilibrium between the *HLA-C* 3′ UTR variant and the *HLA-C* coding regions. Viral load measurements used in categorical analyses were obtained from participants of the Multicentre AIDS cohort study (MACS)[20], the Swiss HIV Cohort (http://www.shcs.ch), the SCOPE cohort[21], and the International HIV Controllers Study Cohort (http://www.hivcontrollers.org). Subjects were classified as HIV controllers if plasma HIV RNA was below 2,000 copies per ml in a minimum of three determinations in the absence of antiretrovirals, spanning at least a 12-month period. Chronically infected individuals were classified as noncontrollers if plasma HIV RNA was above 10,000 copies per ml in the absence of antiretrovirals. The respective institutional review boards approved the study, and all subjects gave written informed consent. All of the patients were of European descent and standard methods for measurements of viral load were used in all three study cohorts.

HLA-C 3′ UTR DNA sequencing. The entire *HLA-C* 3′ UTR was amplified from genomic DNA by PCR using the following pair of primers: forward 5′-gtgag attctggggagctga and reverse 5′-gaacagcaactaggcacagg. The amplicons were sequenced in both directions using the same primers by capillary electrophoresis using an ABI 3730 DNA analyser (Applied Biosystems).

HLA genotyping. DNA samples were genotyped for *HLA-A*, *B* and *C* genes by sequence-based typing of exons 2 and 3 and/or the PCR-sequence-specific oligonucleotide probe typing protocol as recommended by the 13th International Histocompatibility Workshop[24].

Construction of HLA-C 3′ UTR luciferase reporters:. The complete 3′ UTR fragments from various *HLA-C* alleles (*Cw*0303*, *Cw*0401*, *Cw*0502*, *Cw*0602*, *Cw*0701*, *Cw*0702*, *Cw*0802*, *Cw*1203*, *Cw*1502*) were amplified from genomic DNA and inserted into the XbaI site downstream of the luciferase gene in the pGL3-control vector (Promega). Mutations at position 307 were introduced using a site-directed mutagenesis kit (Stratagene) and the following sets of primers: for *07-307C>T* (sense, 5′-gcaacttctacttccctaatgaagttaag<u>aatctg</u>aatataaat ttgtgttc; and antisense, 5′-gaacacaaatttatatt cagattcttaacttcattagggaagtaagaagttgc) and for *07-307C>G* (sense, 5′-caacttcttacttccctaatgaa tta<u>agaagctg</u>aatataaatttgtgtt; and antisense, 5′-aacacaaatttatattcagcttcttaacttcattagggaagtaagaagttg). The intact miR-148-binding site was reconstituted into the *Cw*0602* 3′ UTR by altering only the motif from positions 256–266 and the construct is referred to as *06mut* (5′-CTTTACG-TGT > 5′-ATTCATGGTGC; see Fig. 2). For the *07mut* construct, the miR-148-binding site of the *Cw*0702* 3′ UTR was disrupted by altering positions 256–266 (5′-ATTCATGGTGC > 5′-CTTTACG-TGT). Both *06mut* and *07mut* 3′ UTR sequences were synthesized by Genscript.

Cell lines and culture conditions. The HLA class I negative B721.221 cell line characterized by complete absence of *HLA-A*, *B* and *C* mRNA transcripts[25] were grown in RPMI 1640 (Gibco) medium with 10% heat-inactivated fetal bovine serum (FBS; Atlanta Biologicals) and Epstein–Barr virus (EBV)-transformed B cell lines derived from peripheral blood lymphocytes of individuals homozygous for *HLA-C* alleles (BLCLs) were grown in RPMI 1640 medium with 15% FBS and 1% penicillin–streptomycin–glutamine (PSG; Gibco) at 37 °C in 5% CO_2. The human embryonic kidney cell line (HEK293T)[26] and a T cell line (Jurkat) were cultured in DMEM (Gibco) supplemented with 10% FBS and 1% PSG.

Cell transfection and luciferase reporter assays. B721.221, BLCLs and Jurkat cells were plated at a density of 1×10^6 cells per well in a 96-well plate. 500 ng per well of the pGL3 and 20 ng per well of *Renilla* reporter constructs were transfected using optimized AMAXA nucleofector. 293T cells were plated at a density of 0.5×10^6 cells per well. 50 ng per well of the pGL3 and 2 ng per well of *Renilla* reporter constructs were transfected using Fugene6 (Roche Applied Bioscience). The transfected B721.221 and BLCLs were incubated for 5 h, Jurkat cells were incubated for 24 h, and 293T cells were incubated for 48 h at 37 °C in a CO_2 incubator. The cells were lysed and the firefly and *Renilla* luciferase activities were measured using the Dual Luciferase Reporter Assay System (Promega) and a multidetection microplate reader (fluoSTAR Omega, BMG LABTECH). Luciferase activity of each reporter construct is calculated as fold change relative to the activity of pGL3-basic construct lacking a promoter as well as *HLA-C* 3′ UTR ((luciferase test / average luciferase basic) × (average *Renilla* basic / *Renilla* test)) as previously described[22]. The luciferase activity is presented as fold change of RLU ± s.e. (RLU, mean ± s.e.). Firefly luciferase activity was normalized relative to the *Renilla* luciferase activity for each transfection. All experiments were performed with six replicates and repeated in at least three independent experiments. Similarly, luciferase assays were carried out by co-transfecting 20 pmol per well of oligonucleotide mimics or inhibitors of miR-148a and miR-148b (Dharmacon) or the negative control along with the reporter plasmids. The negative control represents a universal control for both inhibitors and mimics that are

based on the sequences of miRNAs in *C. elegans*. These negative control miRNAs have been confirmed to have minimal sequence identity with miRNAs in human, mouse and rat.

Antibodies and flow cytometry. Before flow cytometry, BLCLs were plated at a density of 1×10^6 cells per well in a 96-well plate and transfected with 20 pmol per well (1 μM final concentration) of either the mimic or inhibitor of miR-148a, miR-148b, or negative control using AMAXA nucleofector (Lonza). The cells were incubated for 24–48 h in a 37 °C CO_2 incubator before determining cell-surface expression for HLA-C or HLA-B.

HLA-C surface expression on BLCLs and peripheral blood lymphocytes was then analysed by staining with the DT9 antibody (provided by V. Braud)[23] or L31 antibody (MediaPharma SRL) followed by a secondary PE-conjugated anti-mouse IgG (Sigma-Aldrich) and collected on an LSRII flow cytometer (BD Biosciences). Isotype controls were obtained from Sigma-Aldrich. HLA-B surface expression on BLCLs was detected using FITC-conjugated anti-Bw4 and anti-Bw6 antibodies (One Lambda) and FITC-conjugated isotype controls. The histograms were plotted using the FlowJo software version 7.5 (TreeStar). L31 binds a linear epitope on HLA class I heavy chain alleles carrying an aromatic residue (Y/F) at position 67 (ref. 27). These include all HLA-C allotypes used in experiments involving L31 in this study and a few crossreacting HLA-B allotypes (HLA B7, B8, B35, B51), which were excluded from samples chosen for our studies involving L31. L31 binds heavy chains that are free of β2m, denatured or unfolded[27].

Lymphocyte separation and miR-148a inhibition. Peripheral blood was obtained from healthy donors and lymphocytes were separated using lymphocyte separation medium as per manufacturer's instructions (Lonza). The lymphocytes were suspended in antibiotic-free RPMI 1640 medium supplemented with 10% FBS and glutamine. PNA inhibitor of miR-148a was added at a final concentration of 0.1 μM for 1×10^6 cells in 200 μl medium and incubated for 48 h. The cells were then analysed for surface expression of HLA-C using the DT9 antibody as described earlier.

Immunoblot analysis. Western blot analysis was performed using protein lysates from 10×10^6 BLCL cell lines transfected with mimic, inhibitor of miR-148a or negative control (1 μM final concentration, transfection conditions as described earlier). The cells were incubated for 48 h before the total cell lysates were prepared. The lysates were subjected to SDS–polyacrylamide (Invitrogen) gel electrophoresis and transferred to Immobilon-P membranes (Millipore). The membranes were blocked at room temperature (20–25 °C) in 5% milk. The membranes were then probed for HLA-C using murine monoclonal antibody L31 (MediaPharma SRL) and β-actin (Santa Cruz Biotechnology). Horseradish peroxidase-conjugated goat anti-mouse IgG (Santa Cruz) was used as the secondary antibody.

Isolation of RNA and miRNA expression analysis by real-time PCR. Total RNA from B721.221 and BLCLs was extracted using the total RNA purification kit (Norgen). Relative quantification of miR-148a, miR-148b, miR-152 and miR-657 was performed using the Taqman real-time PCR assay (Applied Biosystems). RNU48, a small nucleolar RNA that shows abundance and relatively stable expression in both the cell lines used in this study, served as an endogenous control.

Statistical analyses. The effect of polymorphisms in the 3′ UTR of *HLA-C* on HIV viral load control was determined by categorical analyses of the comparison groups: HIV infected controllers versus noncontrollers. Stepwise logistic regression was used to test whether any *HLA* (*HLA-A*, *B* and *C*) allele with a frequency of 1% or greater had an effect in addition to the *HLA-C* 3′ UTR 263 effect. We used SAS procedure PROC LOGISTIC with STEPWISE selection. Sixty-three *HLA* alleles were included in the model. A likelihood ratio test was used to select alleles into the regression model. The significance level for entry of a variable into the model was $P < 0.05$. The *HLA-C* 3′ UTR 263del/del versus 263ins/ins comparison and five *HLA* alleles met the significance level and were selected into the model. SAS9.1 (SAS Institute) was used for data management and statistical analyses. PROC FREQ was used to compute frequencies in each test group. PROC LOGISTIC was used to calculate odds ratios and 95% confidence intervals. A two sided *P* value of < 0.01 was considered statistically significant. PROC ALLELE was used to compute linkage disequilibrium between the 263del/ins SNP and *HLA-C/HLA-B* alleles.

24. Tilanus, M. G. J. in *Immunobiology of the Human MHC. Proceedings of the 13th International Histocompatibility Workshop and Conference* (ed. Hansen, J.) Vol. 1, 304–416 (IHWG Press, 2002).
25. Shimizu, Y. & DeMars, R. Production of human cells expressing individual transferred *HLA-A,-B,-C* genes using an *HLA-A,-B,-C* null human cell line. *J. Immunol.* **142,** 3320–3328 (1989).
26. Graham, F. L., Smiley, J., Russell, W. C. & Nairn, R. Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J. Gen. Virol.* **36,** 59–72 (1977).
27. Setini, A. *et al.* Distinctive features of the α$_1$-domain alpha helix of HLA-C heavy chains free of β$_2$-microglobulin. *Hum. Immunol.* **46,** 69–81 (1996).

LETTER

A system for the continuous directed evolution of biomolecules

Kevin M. Esvelt[1], Jacob C. Carlson[2] & David R. Liu[2,3]

Laboratory evolution has generated many biomolecules with desired properties, but a single round of mutation, gene expression, screening or selection, and replication typically requires days or longer with frequent human intervention[1]. Because evolutionary success is dependent on the total number of rounds performed[2], a means of performing laboratory evolution continuously and rapidly could dramatically enhance its effectiveness[3]. Although researchers have accelerated individual steps in the evolutionary cycle[4–9], the only previous example of continuous directed evolution was the landmark study of Wright and Joyce[10], who continuously evolved RNA ligase ribozymes with an *in vitro* replication cycle that unfortunately cannot be easily adapted to other biomolecules. Here we describe a system that enables the continuous directed evolution of gene-encoded molecules that can be linked to protein production in *Escherichia coli*. During phage-assisted continuous evolution (PACE), evolving genes are transferred from host cell to host cell through a modified bacteriophage life cycle in a manner that is dependent on the activity of interest. Dozens of rounds of evolution can occur in a single day of PACE without human intervention. Using PACE, we evolved T7 RNA polymerase (RNAP) variants that recognize a distinct promoter, initiate transcripts with ATP instead of GTP, and initiate transcripts with CTP. In one example, PACE executed 200 rounds of protein evolution over the course of 8 days. Starting from undetectable activity levels in two of these cases, enzymes with each of the three target activities emerged in less than 1 week of PACE. In all three cases, PACE-evolved polymerase activities exceeded or were comparable to that of the wild-type T7 RNAP on its wild-type promoter, representing improvements of up to several hundred-fold. By greatly accelerating laboratory evolution, PACE may provide solutions to otherwise intractable directed evolution problems and address novel questions about molecular evolution.

We devised a system that exploits the continuous culture and selection of the M13 filamentous bacteriophage[11] (commonly used in phage display[12]) to enable the continuous directed evolution of proteins or nucleic acids. In PACE, *E. coli* host cells continuously flow through a fixed-volume vessel (the 'lagoon') containing a replicating population of phage DNA vectors ('selection phage') encoding the gene(s) of interest (Supplementary Fig. 1).

The average residence time of host cells in the lagoon is less than the time required for *E. coli* replication. As a result, mutations accumulate only in the evolving selection phage population, the only DNA that can replicate faster than the rate of lagoon dilution. The mutation of host cells in the lagoon should therefore have minimal impact on the outcome of the selection over many rounds of phage replication, and mutagenesis conditions are not limited to those that preserve *E. coli* viability.

PACE achieves continuous selection by linking the desired activity to the production of infectious progeny phage containing the evolving gene(s). Phage infection requires protein III (pIII; encoded by gene III), which mediates F pilus binding and host cell entry[13]. Phage lacking pIII

are approximately 10[8]-fold less infectious than wild-type phage[14]. Crucially, the production of infectious phage scales with increasing levels of pIII over concentrations spanning two orders of magnitude[15].

To couple pIII production to the activity of interest, we deleted gene III from the phage vector and inserted it into an 'accessory plasmid' present in the *E. coli* host cells (see Supplementary Fig. 2 for plasmid maps). The production of pIII from the accessory plasmid is dependent on the activity of the evolving gene(s) on the selection phage. Only phage vectors able to induce sufficient pIII production from the accessory plasmid will propagate and persist in the lagoon (Fig. 1). Because pIII expression level determines the rate of infectious phage production[15], phage encoding genes that result in a higher level of pIII production will infect more host cells than phage encoding less active genes.

Owing to the speed of the phage life cycle (progeny phage production begins approximately 10 min after infection)[16], PACE can mediate many generations of selective phage replication in a single day. We observed activity-dependent phage vectors that tolerate lagoon flow rates up to 3.2 volumes per hour (Supplementary Fig. 3), corresponding to an average of 38 phage generations per 24 h (see the Supplementary Information for an analysis). More conservative flow rates of 2.0–2.5 volumes per hour allow 24–30 generations per day and reduce the risk of complete phage loss (washout) during selections. Multiple lagoons can evolve genes in parallel, with each 100 ml lagoon containing approximately 5×10^{10} host cells selectively replicating active phage variants. Importantly, PACE requires no intervention during evolution and obviates the need to create DNA libraries, transform cells, extract genes or perform DNA cloning steps during each round.

In principle, PACE is capable of evolving any gene that can be linked to pIII production in *E. coli*. Because a wide variety of functions including DNA binding, RNA binding, protein binding, bond-forming catalysis and a variety of enzyme activities have been linked to the expression of a reporter protein[17,18], PACE can be applied to the evolution of many different activities of interest. As examples, we successfully linked protein–protein binding, recombinase activity and RNAP activity to phage infectivity in discrete infection assays by creating variants of the accessory plasmid that associate each of these activities with pIII production (Fig. 2).

PACE applies optimal evolutionary pressure when pIII levels are above the minimal threshold required to prevent phage washout, but below the amount needed to maximize infectious phage production. This window can be shifted by varying the copy number of the accessory plasmid, or by altering the ribosome-binding site (RBS) sequence of gene III to modulate the efficiency with which gene III is expressed (Supplementary Fig. 4).

We constructed an arabinose-inducible mutagenesis plasmid that elevates the error rate during DNA replication in the lagoon by suppressing proofreading[19] and enhancing error-prone lesion bypass (Supplementary Information)[20]. Full induction increased the observed mutagenesis rate by approximately 100-fold, inducing all possible transitions and transversions (Supplementary Fig. 5). This enhanced

[1]Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [2]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA. [3]Howard Hughes Medical Institute, Cambridge, Massachusetts 02138, USA.
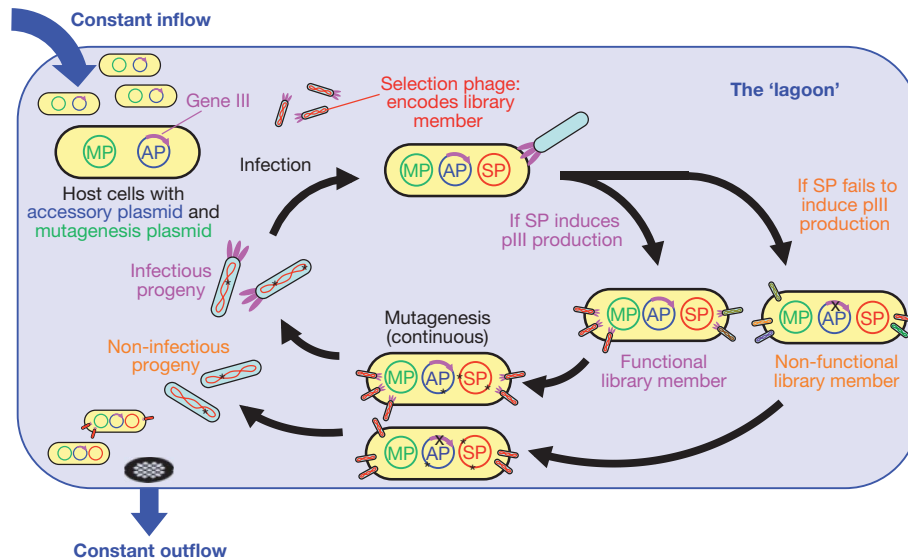
Figure 1 | Overview of the PACE system. PACE in a single lagoon. Host cells continuously flow through a lagoon, where they are infected with selection phage (SP) encoding library members. Functional library members induce production of pIII from the accessory plasmid (AP) and release progeny capable of infecting new host cells, whereas non-functional library members do not. Increased mutagenesis is triggered through induction of the mutagenesis plasmid (MP). Host cells flow out of the lagoon on average faster than they can replicate, confining the accumulation of mutations to replicating phage.

mutation rate is sufficient to sample all possible single and double mutants of a given phage-encoded sequence in each generation (Supplementary Information), in principle enabling single-mutation fitness valleys to be traversed during PACE.

Bacteriophage T7 RNAP is widely used to transcribe RNA *in vitro* and in cells. T7 RNAP is highly specific for its promoter sequence (TAATACGACTCACTATA), and exhibits virtually no detectable activity on the consensus promoter of the related bacteriophage T3 (AATTAACCCTCACTAAA, differences underlined)[21,22]. Despite decades of study and several attempts to engineer the specificity of T7 RNAP towards other promoters[22,23] including that of T3, a mutant T7 RNAP capable of recognizing the T3 promoter has not been previously reported.

To remove potential interference from evolutionary improvements to the phage vector rather than to T7 RNAP, we propagated a selection phage expressing wild-type T7 RNAP for 3 days on host cells containing an accessory plasmid with the wild-type T7 promoter driving gIII expression. A single plaque presumed to represent vector-optimized
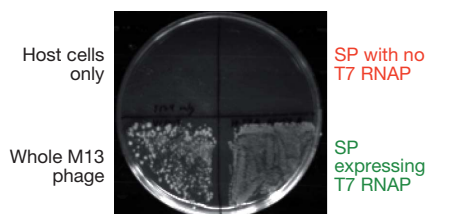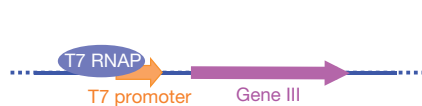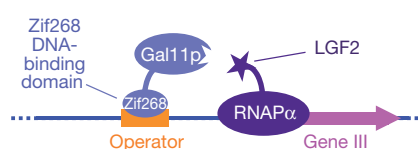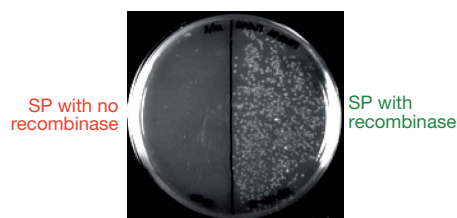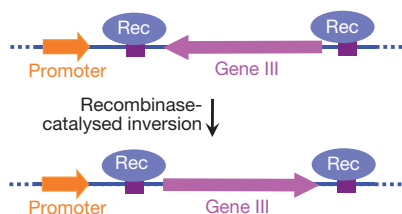
Figure 2 | Linkage of three protein activities to pIII production and phage infectivity using three distinct accessory plasmids. *E. coli* cells containing accessory plasmids encoding conditionally expressed gene III (left) and selection phage were combined with recipient cells. Phage production resulted in colonies with antibiotic resistances conferred by the phage and the recipient cells (right). See Methods for details. **a**, RNAP activity leads to gene III expression and infection comparable to wild-type phage, whereas selection phage lacking T7 RNAP do not infect. **b**, Protein–protein interaction between a Gal11p domain tethered to a Zif268 DNA-binding domain and an LGF2a domain fused to RNAP leads to gene III expression and infection. **c**, Recombinase-catalysed gene inversion induces gene III expression and infection.

selection phage contained a single mutation (P314T) in T7 RNAP. We confirmed that the activity of the P314T mutant does not significantly differ from that of wild-type T7 RNAP (Supplementary Fig. 6).

This starting selection phage failed to propagate on host cells containing the T3 promoter accessory plasmid. We therefore propagated the phage on cells containing a hybrid T7/T3 promoter accessory plasmid with the T7 promoter base at the important −11 position[21] but all other positions changed to their T3 counterparts. Two initially identical lagoons were evolved in parallel on the hybrid promoter accessory plasmid for 60 h, then on the complete T3 promoter accessory plasmid for 48 h, and finally on a high-stringency, very low-copy T3 promoter accessory plasmid for 84 h (Fig. 3a).

Phage persisted in both lagoons after 8 days of PACE, surviving a net dilution of 10^{167}-fold, the equivalent of 555 phage population doublings and 200 rounds of evolution by the average phage. We isolated, sequenced and characterized phage vectors from each lagoon after 48, 108 and 192 h, observing up to 8, 10 and 11 non-silent mutations in single T7 RNAP genes at each time point.

Protein-encoding regions (without upstream promoter sequences) of evolved mutant T7 RNAP genes were subcloned into assay plasmids that quantitatively link transcriptional activity to β-galactosidase expression in cells[24]. We defined the activity of wild-type T7 RNAP

on the T7 promoter to be 100%. The starting T7 RNAP exhibited undetectable (<3%) levels of activity on the T3 promoter in these cell-based assays. The assayed mutants exhibited more than 200% activity after 108 h of PACE, and more than 600% activity after high-stringency PACE at 192 h, improvements of more than 200-fold (Fig. 3b, c). These results collectively establish the ability of PACE to evolve large changes in enzyme activity and specificity very rapidly with minimal intervention by the researcher.

Several evolved T7 RNAP mutants were also purified and assayed *in vitro* using radioactive nucleotide incorporation assays. Purified T7 RNAP mutants exhibited activity levels on the T3 promoter *in vitro* exceeding that of wild-type T7 RNAP on the T7 promoter, representing improvements of up to 89-fold compared with the starting enzyme (Supplementary Fig. 7). These results indicate that PACE resulted in large improvements in substrate binding or catalytic rate. Evolved activity improvements were higher in cells than *in vitro*, suggesting that these enzymes also evolved improvements in features such as expression level, polymerase folding or stability that are specific to the context of the cytoplasm.

Interestingly, the evolutionary dynamics of the two initially identical lagoons differed significantly (Fig. 3d and Supplementary Results). Within 24 h, lagoon 1 acquired a predominant suite of mutations
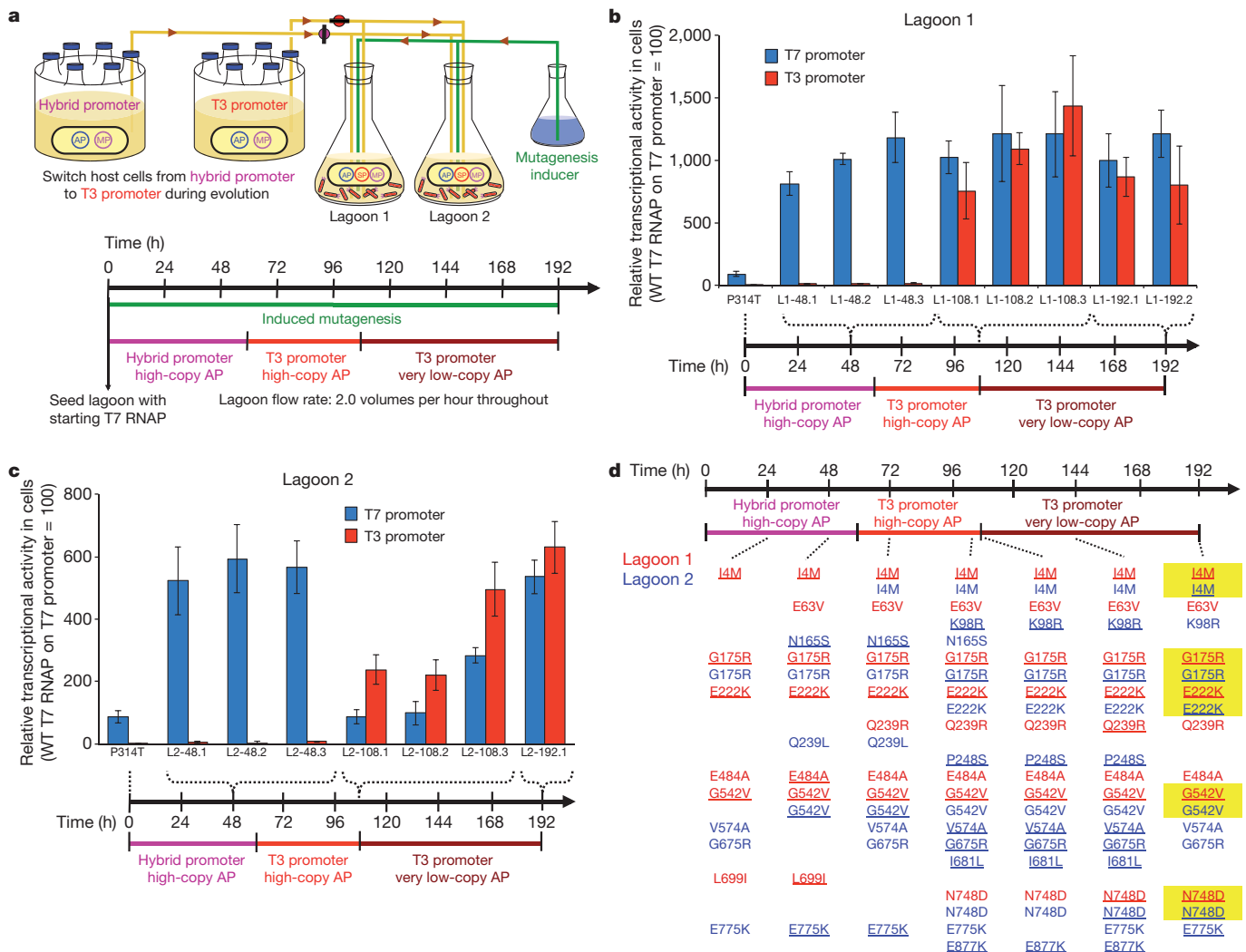
Figure 3 | Continuous evolution of T7 RNAP variants that recognize the T3 promoter. a, PACE schedule. **b**, Activity in cells of T7 RNAP variants isolated from lagoon 1 at 48, 108 and 192 h on the T7 and T3 promoters. Transcriptional activity was measured spectrophotometrically by subcloning the protein-encoding regions of the T7 RNAP genes into a construct in which the T7 or T3 promoter drives *lacZ* expression. **c**, Activity in cells of T7 RNAP

variants isolated from lagoon 2. Error bars in **b** and **c**, standard deviation of at least three independent assays. **d**, Mutations identified in T7 RNAP clones from lagoons 1 and 2 are shown in red and blue, respectively. Underlined mutations were predominant in that lagoon based on whole-pool sequencing of lagoon aliquots. Mutations highlighted in yellow independently evolved to predominance in both lagoons.

consisting of I4M, G175R, E222K and G542V and changed thereafter beyond acquiring N748D, a mutation known to enable recognition of the T3 base at the −11 position[21], after exposure to the full T3 promoter. In contrast, lagoon 2 accessed these mutations more slowly before a different suite of mutations also including N748D became predominant at 108 h, only to be displaced by the same suite of mutations observed in lagoon 1. The presence of several mutations unique to lagoon 2 throughout the experiment suggests that lagoon cross-contamination did not occur. The distinct evolutionary trajectories of the two lagoons before their ultimate convergence upon a common set of mutations highlight the ability of PACE to discover multiple viable pathways to a target activity in parallel experiments. This capability may enable a more in-depth experimental study of protein evolutionary dynamics than can be achieved with conventional directed evolution methods that cannot complete so many rounds of evolution on a practical timescale.

T7 RNAP is highly specific for initiation with GTP[25,26], significantly limiting its usefulness for the *in vitro* transcription of RNAs that begin with other nucleotides. As initiation has been described as a mechanistically challenging step in transcription[27], we next used PACE to evolve T7 RNAP variants capable of initiating transcription with other nucleotides in a template-directed manner. T7 RNAP is known to initiate preferentially with GTP up to several bases downstream of the +1 position if the template is devoid of early guanines in the coding strand[26]. We therefore constructed accessory plasmids in which positions +1 to +6 of the gene III transcript were AAAAAA (iA_6) or CCCCCC (iC_6).

We used PACE to rapidly evolve variants of T7 RNAP capable of initiating with ATP. In view of previous reports indicating varying degrees of initiation of T7 RNAP with ATP[25,26], we propagated starting phage in host cells with a high-copy iA_6 accessory plasmid for 24 h, followed by a 30:70 high-copy:very-low-copy mixture of host cells for 12 h (Fig. 4a). At a dilution rate of 2.5 volumes per hour, phage survived a total dilution of 10^{39}-fold and experienced an average of approximately 45 rounds of evolution.

The wild-type enzyme exhibited undetectable initial activity in cells (<3%) on the iA_6 promoter. All six clones isolated after only 36 h of PACE exhibited at least 170% activity on the iA_6 promoter in cell-based assays, while retaining at least 120% activity on the wild-type promoter (Fig. 4b). Purified variants assayed *in vitro* exhibited activities on the iA_6 promoter matching that of wild-type T7 RNAP on the T7 promoter (Supplementary Fig. 8). RACE analysis of transcripts produced by the most active clone (A6-36.4) confirmed that this enzyme begins transcripts with the template-directed bases on the iA_6, iC_6 and wild-type promoters (Supplementary Fig. 9a). All six characterized clones contained K93T, S397R and S684Y mutations, whereas three of the six also contained S228A (Supplementary Table 2). Residue 397 directly contacts the nascent RNA strand[28], suggesting a role for S397R in allowing efficient initiation of iA_6 transcripts.

We concurrently evolved T7 RNAP to initiate transcripts with CTP (Supplementary Fig. 10a). We observed that wild-type T7 RNAP retains significant activity on the iC_6 promoter (~50%) both in cells and *in vitro* (Supplementary Fig. 10b, c), a surprising observation in view of reports that the enzyme initiates with G at the +2 position if the +1 position is C[25]. Although the high starting activity precluded large improvements, the most active PACE-evolved variants nevertheless exceeded 100% activity on the iC_6 promoter both in cells and *in vitro* (Supplementary Fig. 10b, c and Supplementary Results). RACE analysis of transcripts produced by the most active clone (C6-80.9) confirmed that this enzyme begins transcripts with the template-directed bases (Supplementary Fig. 9b).

The three PACE experiments executed 45–200 rounds of evolution in 1.5–8 days and yielded T7 RNAP variants with activities on their target promoters or templates that exceeded or matched the activity of the wild-type enzyme transcribing the wild-type T7 promoter both in cells and *in vitro*. This degree of improvement is especially significant
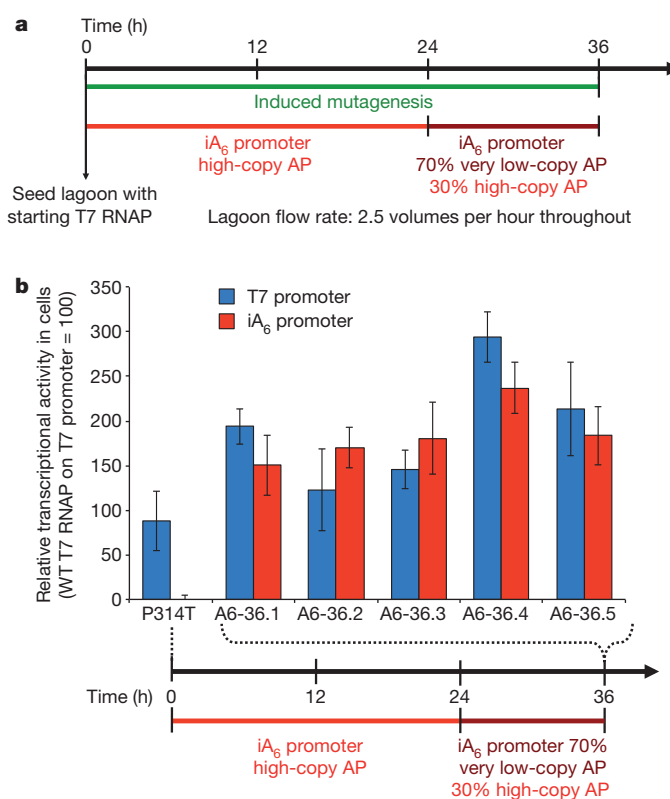
Figure 4 | Continuous evolution of T7 RNAP variants that initiate transcription with A. **a**, PACE schedule. **b**, Activity in cells of T7 RNAP variants on the T7 and iA_6 promoters isolated after 36 h of PACE. Assays were performed as described in Fig. 3b. Error bars, standard deviation of at least three independent assays.

given that, for two of the evolved activities, the starting polymerase exhibited virtually no detectable activity.

The evolved A6-36.4 variant of T7 RNAP can initiate transcription from iC_6, iA_6 and wild-type templates in a template-directed manner with efficiencies comparable to that of wild-type T7 RNAP initiating with the wild-type template (Supplementary Fig. 11) and sequence fidelity sufficient to mediate the production of functional pIII and LacZ enzyme. These findings suggest that this enzyme, and possibly other PACE-evolved variants, may represent improved, more general T7 RNAPs for routine *in vitro* and *in vivo* transcription.

The PACE system can be assembled entirely from a modest collection of commercially available equipment (listed in Supplementary Table 3) and does not require the manufacture of any specialized components. The ability to perform dozens of rounds of evolution each day with minimal researcher involvement implies that PACE is particularly well suited to address problems or questions in molecular evolution that require hundreds to thousands of generations, or the execution of many evolution experiments in parallel. More generally, PACE represents the integration and manipulation of many protein and nucleic acid components in a living system to enable the rapid generation of biomolecules with new activities, a significant example and goal of synthetic biology.

METHODS SUMMARY

Discrete infection assays. Donor culture (2 µl) containing accessory plasmid and selection phage were mixed with 198 µl of F+ recipient cells and incubated for 1.5 h at 37 °C. Dilutions were plated to select for recipient cells containing selection phage. **PACE.** *E. coli* host cells with accessory plasmid and mutagenesis plasmid (see Methods) maintained at 5×10^8 cells ml^{-1} were pumped at 2.0, 2.5 or 3.2 volumes per hour into a fixed-volume lagoon seeded with selection phage. Aliquots were taken regularly. Evolved mutants were isolated as individual plaques and subcloned for cell-based or *in vitro* activity assays.

Transcription assays. Activity in cells was measured using a standard β-galactosidase expression assay[24]. *In vitro* activity was assessed using a standard radioactive nucleotide incorporation assay[29] with His-tag-purified enzyme variants.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Yuan, L., Kurek, I., English, J. & Keenan, R. Laboratory-directed protein evolution. *Microbiol. Mol. Biol. Rev* **69,** 373–392 (2005).
2. Voigt, C. A., Kauffman, S. & Wang, Z. G. Rational evolutionary design: the theory of *in vitro* protein evolution. *Adv. Protein Chem.* **55,** 79–160 (2000).
3. Mills, D. R., Peterson, R. L. & Spiegelman, S. An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proc. Natl Acad. Sci. USA* **58,** 217–224 (1967).
4. Wang, L., Jackson, W. C., Steinbach, P. A. & Tsien, R. Y. Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proc. Natl Acad. Sci. USA* **101,** 16745–16749 (2004).
5. Camps, M., Naukkarinen, J., Johnson, B. P. & Loeb, L. A. Targeted gene evolution in *Escherichia coli* using a highly error-prone DNA polymerase I. *Proc. Natl Acad. Sci. USA* **100,** 9727–9732 (2003).
6. Makeyev, E. V. & Bamford, D. H. Evolutionary potential of an RNA virus. *J. Virol.* **78,** 2114–2120 (2004).
7. Davis, J. N. & van den Pol, A. N. Viral mutagenesis as a means for generating novel proteins. *J. Virol.* **84,** 1625–1630 (2009).
8. Das, A. T. *et al.* Viral evolution as a tool to improve the tetracycline-regulated gene expression system. *J. Biol. Chem.* **279,** 18776–18782 (2004).
9. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460,** 894–898 (2009).
10. Wright, M. C. & Joyce, G. F. Continuous *in vitro* evolution of catalytic function. *Science* **276,** 614–617 (1997).
11. Husimi, Y. Selection and evolution of bacteriophages in cellstat. *Adv. Biophys.* **25,** 1–43 (1989).
12. Smith, G. P. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228,** 1315–1317 (1985).
13. Riechmann, L. & Holliger, P. The C-terminal domain of TolA is the coreceptor for filamentous phage infection of *E. coli*. *Cell* **90,** 351–360 (1997).
14. Nelson, F. K., Friedman, S. M. & Smith, G. P. Filamentous phage DNA cloning vectors: a noninfective mutant with a nonpolar deletion in gene III. *Virology* **108,** 338–350 (1981).
15. Rakonjac, J. & Model, P. Roles of pIII in filamentous phage assembly. *J. Mol. Biol.* **282,** 25–41 (1998).
16. Calendar, R. *The Bacteriophages* (Oxford Univ. Press, 2006).
17. Vidal, M. & Legrain, P. Yeast forward and reverse 'n'-hybrid systems. *Nucleic Acids Res.* **27,** 919–929 (1999).
18. Baker, K. *et al.* Chemical complementation: a reaction-independent genetic assay for enzyme catalysis. *Proc. Natl Acad. Sci. USA* **99,** 16537–16542 (2002).
19. Fijalkowska, I. J. & Schaaper, R. M. Mutants in the Exo I motif of *Escherichia coli* dnaQ: defective proofreading and inviability due to error catastrophe. *Proc. Natl Acad. Sci. USA* **93,** 2856–2861 (1996).
20. Opperman, T., Murli, S., Smith, B. T. & Walker, G. C. A model for a umuDC-dependent prokaryotic DNA damage checkpoint. *Proc. Natl Acad. Sci. USA* **96,** 9218–9223 (1999).
21. Raskin, C. A., Diaz, G., Joho, K. & McAllister, W. T. Substitution of a single bacteriophage T3 residue in bacteriophage T7 RNA polymerase at position 748 results in a switch in promoter specificity. *J. Mol. Biol.* **228,** 506–515 (1992).
22. Ikeda, R. A., Chang, L. L. & Warshamana, G. S. Selection and characterization of a mutant T7 RNA polymerase that recognizes an expanded range of T7 promoter-like sequences. *Biochemistry* **32,** 9115–9124 (1993).
23. Raskin, C. A., Diaz, G. A. & McAllister, W. T. T7 RNA polymerase mutants with altered promoter specificities. *Proc. Natl Acad. Sci. USA* **90,** 3147–3151 (1993).
24. Vidal-Aroca, F. *et al.* One-step high-throughput assay for quantitative detection of beta-galactosidase activity in intact Gram-negative bacteria, yeast, and mammalian cells. *Biotechniques* **40,** 433–438 (2006).
25. Imburgio, D., Rong, M., Ma, K. & McAllister, W. T. Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. *Biochemistry* **39,** 10419–10430 (2000).
26. Brieba, L. G., Padilla, R. & Sousa, R. Role of T7 RNA polymerase His784 in start site selection and initial transcription. *Biochemistry* **41,** 5144–5149 (2002).
27. Kuzmine, I., Gottlieb, P. A. & Martin, C. T. Binding of the priming nucleotide in the initiation of transcription by T7 RNA polymerase. *J. Biol. Chem.* **278,** 2819–2823 (2003).
28. Cheetham, G. M., Jeruzalmi, D. & Steitz, T. A. Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature* **399,** 80–83 (1999).
29. Martin, C. T. & Coleman, J. E. Kinetic analysis of T7 RNA polymerase-promoter interactions with small synthetic promoters. *Biochemistry* **26,** 2690–2696 (1987).

METHODS

General methods. All equipment, reagents, suppliers and relevant catalogue numbers are detailed in Supplementary Table 3. All PCR reactions were performed with HotStart Phusion II polymerase. Water was purified using a MilliQ water purification system (Millipore).

DNA cloning. All vectors were constructed by isothermal assembly cloning.[30] Isothermal assembly buffer (5×) contained 3 ml 1 M Tris-HCl pH 7.5, 300 μl 1 M MgCl$_2$, 600 μl 10 mM deoxynucleotide triphosphates, 300 μl 1 M dithiothreitol, 1.5 g PEG-8000, 20 mg NAD, and H$_2$O to 6 ml. Individual 320 μl aliquots were frozen at −20 °C. Isothermal assembly master mix was prepared by mixing 320 μl 5× buffer with 1 μl T5 exonuclease, 20 μl Phusion polymerase, 160 μl Taq DNA ligase, and H$_2$O to 700 μl. Individual 15 μl aliquots in PCR tubes were frozen at −20 °C. DNA fragments to be assembled were PCR-amplified using oligonucleotide primers designed to ensure between 30 and 40 base pairs (bp) of overlap homology with each adjacent fragment. DpnI was added directly to the PCR reactions to remove template DNA, followed by PCR cleanup with MinElute columns according to the manufacturer's protocol. Fragments were assembled by mixing equimolar amounts totalling 5 μl with 15 μl isothermal assembly master mix and incubating at 50 °C for 1 h. Assembly mixtures were directly transformed into NEB Turbo competent cells by heat shock or purified by MinElute columns as described before electroporation.

Plasmids. T7 RNAP-dependent accessory plasmids contained, in order, a strong *rrnB* terminator, the promoter of interest, a desired RBS, gene III, the *bla* gene conferring carbenicillin resistance, and either the pUC or SC101 origin of replication. For selection stringency assays, RBS A = 5′-AAGGAGGTAACTCATAG TG-3′, RBS B = 5′-AAGGAAATAACTCATAGTG-3′ and RBS C = 5′-AAG AAAATAACTCATAGTG-3′, where underlined bases represent the start codon of gene III. Reporter plasmids were identical to SC101 accessory plasmids except for the replacement of gene III by full-length *lacZ*. T7 RNAP selection phage was constructed by replacing all but the last 180 bp of gene III with the gene encoding T7 RNAP in VCSM13 helper phage. The mutagenesis plasmid consisted of dnaQ926, umuD′, umuC and recA730 under control of the araC operon. Expression plasmids used for quantification assays consisted of the cloD13 origin of replication, aadA and the wild-type gene III promoter and RBS driving expression of the evolved T7 RNAP variant. All plasmids used in this work are described in Supplementary Table 1. Vector maps of representative plasmids are shown in Supplementary Fig. 2.

Bacterial strains. All DNA cloning was performed with Mach1 cells or NEB Turbo cells. Early discrete infection assays and PACE experiments were performed with PirPlus DH10βF′DOT cells. Plaque assays and PACE experiments with T7 RNAP were performed with using *E. coli* S109 cells derived from DH10B by replacement of the proBA locus with the *pir116* allele, as previously described.[31] To our knowledge, this modification was not required for PACE experiments with T7 RNAP. Similarly, the *lacI* cassette was deleted from the F plasmid and from the chromosome to enable mutagenesis assays. S109 cells were rendered F$^+$ by conjugation with ER2738. The complete genotype of the resulting strain is F′proA+B+ Δ(lacIZY) zzf::Tn10(TetR)/ endA1 recA1 galE15 galK16 nupGrpsLΔlacIZYA araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) proBA::pir116 λ⁻.

Discrete infection assays. Discrete (non-continuous) infectivity assays were performed by co-transforming phage bearing antibiotic resistance genes with the appropriate accessory plasmid into competent cells to generate phage donors. For two-hybrid and recombinase experiments, phage-producing cells contained gIII-deleted helper phage, accessory plasmid and selection phagemid. The recombinase used was Hin, a member of the serine resolvase family. Colonies were picked and grown overnight in 2× YT media with both antibiotics. Donor cells (2 μl) were mixed with 198 μL F$^+$ recipient cells in mid-exponential phase containing an antibiotic resistance gene not found in the donor. Mixtures were incubated at 37 °C for 1.5 h and 20 μl of serial dilutions were spread on plates containing the donor and the recipient antibiotics. Infection was quantified by the number of resulting colonies after incubation at 37 °C overnight. For plaque assays, phage DNA was transformed into electrocompetent cells containing the appropriate accessory plasmid and recovered for 1 h at 37 °C, or simply isolated from a lagoon. Serial dilutions were mixed with 300 μl F$^+$ recipient cells grown to exponential phase in 14 ml Falcon culture tubes and incubated at 37 °C for 15 min. Top agar (3 ml) (7 g l^{-1} from Luria–Bertani (LB) broth base) at 50 °C was added to each tube, briefly vortexed and poured onto minimal agar plates incubated at 37 °C for 8 h or overnight to generate plaques.

Turbidostat assembly. Assembly followed the schematic shown in Supplementary Fig. 1c. Turbidostats were constructed from BioProbe flasks on magnetic stir plates. Each flask was equipped with a TruCell2 cell density meter held in a GL32 probe holder with compression fitting. GL45 and GL32 septa pierced with needles transferred media to and from the turbidostat through an 8-channel peristaltic pump with Tygon tubing. A needle set at the desired turbidostat volume

level pumped excess cells to the waste container. A 0.2 μm filter attached to a 14-gauge needle piercing the septum vented the turbidostat vessel. A two-way valve controlled media flow to the turbidostat, connected such that a closed valve state returned the media to source. The valve opened and closed in response to TruCell2 4–20 mA output processed by a digital panel meter programmed with the desired set point. Panel meters were unlocked according to the instruction manual and programmed to the following settings: Input dc_A, Setup 30_ 10, Config 00000, Filtr 11009, dec.pt ddd.dd, lo in 00.400, lo rd 004.00, hi in 02.000, hi rd 020.00, Alset 00036, deu1h 000.01, deu2h 000.01. The digital panel meter and valve were connected with a solderless breadboard according to the diagram shown in Supplementary Fig. 12. Lagoons consisted of 100 ml Pyrex bottles with GL45 septa pierced with needles for fluid delivery, a 0.2 μm filter-terminated vent line and a magnetic stir bar. Excess lagoon volume was continually pumped to waste through a waste needle set at the desired lagoon volume.

Media preparation. Each 20 l media carboy received 140 g anhydrous potassium phosphate dibasic, 40 g potassium phosphate monobasic, 20 g ammonium sulphate, and 20 ml Tween 80 in 20 l H$_2$O. Carboys were loosely capped with Polyvent filling/ venting closures with an autoclavable 0.2 μm filter fastened to the venting port. Media was autoclaved until visibly boiling (typically 120 min at 30 pounds per square inch (p.s.i.), 121 °C) and allowed to cool overnight. Media supplement was prepared from 90 g glucose, 10 g sodium citrate, 0.25 g anhydrous magnesium sulphate, 10 g casamino acids, 0.15 g tetracycline-HCl, 0.6 g carbenicillin, 0.6 g spectinomycin and 0.5 g (L)-leucine, dissolved in 500 ml H$_2$O, and filtered with a Nalgene 500 ml filtration unit. Media supplement (500 ml) was added to each carboy under conditions that minimized the risk of media contamination (in the case of the reported experiments, immediately after 1 h of germicidal ultraviolet irradiation).

Sterilization and cell culture. The autoclavable components of a turbidostat apparatus include the BioProbe flask, TruCell2 probe, needles, vent filter and tubing. All such components were autoclaved fully assembled except for tubing, which was connected while hot. Upon equilibrating to ambient temperature, the peristaltic pump responsible for media addition and waste removal was started with the valve opened until the desired volume was reached. The TruCell2 probe was connected to its transmitter and zeroed. Turbidostats were seeded with 100 μl of an overnight culture of host cells. Turbidostats and lagoon cultures were grown at 37 °C.

Cell density calibration. Serial dilution plating was used to generate a calibration curve to determine TruCell2 output and panel meter setting corresponding to the desired cell density. For these experiments, both panel meter alarms were programmed to open the valve at 6.80 mA. Cells were pumped from the turbidostat to the lagoons by peristaltic pumps with silicone (platinum) two-stop tubing. Calibration curves relating pump speed in revolutions per minute to volumetric flow rate were determined experimentally with a timer and graduated cylinder for each tubing size.

PACE experiments. Turbidostats and lagoons were assembled as described above. Upon the turbidostat reaching the desired set point of 6.80 mA (corresponding to 5×10^8 cells ml^{-1} in our hands), lagoons were connected to turbidostats, waste needles were set at the desired volume and lagoon pumps were set to a flow rate corresponding to the desired dilution rate. Each lagoon was seeded with 100 μl of an overnight culture producing selection phage. To induce elevated mutagenesis, 10% filter-sterilized (L)-arabinose was delivered by a separate peristaltic pump to each lagoon requiring enhanced mutagenesis to a final concentration of 1%. Lagoon aliquots were taken by sampling lagoon waste lines at the luer lock just after the peristaltic pump. Individual clones were isolated by plaque assay or amplified by PCR, assembled into a T7 RNAP activity assay plasmid, and transformed into cells containing a *lacZ* reporter plasmid. Active clones were picked by blue/white screening.

Selection phage optimization. T7 RNAP was subcloned into VCSM13 helper phage encoding kanamycin resistance, generating HP-T7RNAP A, used in the discrete infection assay shown in Fig. 2a. To ensure that improvements to the phage genome did not interfere with the evolution of T7 RNAP, HP-T7RNAP A was propagated in a lagoon fed by S109 host cells containing AP-T7 A and DP-QUR and supplemented with arabinose for 72 h at 2.0 volumes per hour. Individual plaques were isolated and their T7 RNAP genes sequenced. One plaque contained only a single point mutation in T7 RNAP, P314T, and was chosen as the preoptimized selection phage for T7 RNAP evolution. Sequencing of the rest of the selection phage revealed numerous changes relative to the parental VCSM13. Notably, the entire p15a-Kan^R cassette inserted into the intergenic region to create VCSM13 had been perfectly deleted to reconstitute the wild-type M13 intergenic region. Other changes included N79S, F286S and I360T mutations in gIV, a K249R mutation in gII, three silent mutations back to the corresponding M13 base, two other silent mutations and the deletion of one thymine residue in the terminator before gIII, possibly increasing the expression of T7 RNAP. These regional patterns of variation parallel those observed by Husimi in more extensive filamentous phage evolution

experiments[11]. This evolved phage, designated SP-T7RNAP P314T, was used as the starting selection phage for all subsequent PACE experiments.

Mutagenesis assays. The *lacI* gene was cloned into VCSM13 between the p15a origin and *kan*[r] to generate VCSM13-lacI. A turbidostat was grown to a set point equivalent to 5×10^8 cells ml^{-1} with S109 cells containing the mutagenesis plasmid. Lagoons were seeded with 10 μl VCSM13-lacI and run at 2.5 volumes per hour for 3 h. One lagoon was supplemented with 10% L-arabinose to a final concentration of 1%, whereas the other received H$_2$O. Aliquots were removed after 3 h and each was used to infect a 100-fold greater volume of recipient cell culture of S109 cells containing a reporter plasmid conferring carbenicillin and spectinomycin resistance with a *lacI* binding site (*lacO*) capable of repressing spectinomycin resistance. Mixtures were incubated for 1.5 h at 37 °C, spread on 2× YT plates containing spectinomycin, kanamycin and carbenicillin, and incubated at 37 °C overnight. Colonies were counted for induced and uninduced lagoons to estimate the fold increase in mutagenesis. Seventy-two colonies were sequenced to determine the frequencies of all transitions and transversions within mutated *lacI* genes. The results are shown in Supplementary Fig. 5. All sequenced colonies contained at least one mutation capable of inactivating repressor function.

Cell-based T7 RNAP activity assays. Overnight cultures of S109 cells grown in 2× YT containing reporter plasmid and expression vector were diluted fourfold in fresh 2× YT media. Diluted culture (20 μl) was mixed with 80 μl Z buffer (60 mM Na$_2$HPO$_4$, 40 mM NaH$_2$PO$_4$, 10 mM KCl, 1 mM MgSO$_4$, 50 mM β-mercaptoethanol, pH 7.0) in Falcon Microtest 96-well OptiLux assay plates and the absorbance at 595 nm was measured using a Spectra M5 plate reader. Methylumbelliferyl-β-(D)-galactopyranoside (25 μl of 1 mg ml^{-1}) was added to each well and the time recorded. Plates were incubated at 30 °C and fluorescence was measured at 360/460 nm on a Spectra M5 plate reader. Plates were measured at multiple time points to avoid saturation of the spectrophotometer or consumption of the substrate, depending on the activity level of the T7 RNAP enzyme being assayed. MUG fluorescence units were calculated as previously described[24]. The activity level of wild-type T7 RNAP on the T7 promoter was defined as 100%; activities greater than 3% were considered significantly above the background level of this assay.

T7 RNAP protein purification. T7 RNAP variants were cloned into pT7-911Q, a His-tagged T7 RNAP expression vector.[32] Overnight cultures grown at 30 °C were diluted 1:500 in LB broth containing 50 μg ml^{-1} carbenicillin and 2% glucose. Upon reaching absorbance $A_{600} \approx 0.5$, cultures were centrifuged at 4,000*g* for 5 min and re-suspended in LB broth with 0.4 mM IPTG and 50 μg ml^{-1} carbenicillin. Cultures were grown for 4 h at 30 °C, spun at 8,000*g* for 6 min, and the pellet was frozen overnight. Binding buffer consisted of 50 mM Tris, 300 mM NaCl, 5% glycerol, 5 mM β-mercaptoethanol and 10 mM imidazole at pH 8.0. Wash buffer consisted of 50 mM Tris, 800 mM NaCl, 5% glycerol, 5 mM β-mercaptoethanol and 20 mM imidazole at pH 8.0. Elution buffer was equivalent to wash buffer with 500 mM imidazole. Pellets from 25 ml culture were resuspended in 1 ml wash buffer and cells were lysed by sonication while kept on ice using a Misonix CL4 sonicator at maximal microtip power for 45 s in 1 s bursts. Cell debris was spun down at 20,000*g* for 15 min at 4 °C. Ni-NTA spin columns were equilibrated with 500 μl binding buffer and spun at 800*g* for 2 min. Lysate supernatant was loaded onto each column and spun at 300*g* for 5 min. Columns were washed twice with 500 μl wash buffer, spinning at 800*g* for each, then eluted twice with 250 μl elution buffer. Proteins were dialysed into 20 mM Tris, 100 mM NaCl, 5% glycerol, 1 mM EDTA, 1 mM DTT, pH 8.0, and concentrated using Amicon Ultra-0.5 30K concentration columns.

***In vitro* T7 RNAP activity assays.** Purified T7 RNAP variant concentrations were determined by Bradford assay and then by Coomassie stain on a 4–12% NuPage gel. Templates were prepared by PCR amplification of 150-bp fragments of the reporter plasmids used for *in vivo* assays including the promoter and the start of the *lacZ* gene. Templates were purified by MinElute spin column. Transcription reactions were performed in 1× RNAP buffer consisting of 40 mM Tris-HCl, 6 mM MgCl$_2$, 10 mM dithiothreitol, 2 mM spermidine pH 7.9 with 1 mM ribonucleotide tri-phosphates, 1 ng template DNA, purified polymerase variant and 2 mCi [α-^{32}P]ATP. Reactions were incubated at 37 °C for 20 min, mixed with an equivalent volume of loading dye consisting of 7 M urea, 178 mM Tris-Cl, 178 mM

H$_3$BO$_3$, 4 mM EDTA and 0.002% bromophenol blue, then electrophoresed on Criterion 5% or 10% TBE-urea denaturing gels. RNAs were transcribed from double-stranded templates of sequence 5′-TAATACGACTCACTATA<u>GGG</u>AGA GCCACCACCACCACCACCACCA-3′, 5′-TAATACGACTCACTATA<u>C</u>CCCC CGCCACCACCACCACCACCACCA-3′, and 5′-TAATACGACTC<u>A</u>CTAT <u>A</u>AAAAAAGCCACCACCACCACCACCACCA-3′ (the +1 base is underlined). To remove differences in specific radioactivity from [α-^{32}P]ATP incorporation in iA$_6$, wild-type and iC$_6$ transcripts arising from the differing number of A nucleotides in their first six bases, transcripts were digested with T1 RNase to remove the first seven nucleotides. This digestion step also enabled all iA$_6$ transcription products to enter the gel completely, which did not always occur for iA$_6$ products of either wild-type or evolved polymerases. Transcripts were electrophoresed on Criterion 15% TBE-urea gels, exposed to phosphor screens, and imaged on a Typhoon Trio phosphorimager. Bands corresponding to transcription products were quantified with ImageJ software.

RACE analysis. *In vitro* transcription reactions with purified T7 RNAP variants were performed as described above but without the addition of any radioactive nucleotide using polymerase variant C6-80.9 on the iC$_6$ and wild-type templates, and using polymerase variant A4-36.4 on the iA$_6$, iC$_6$ and wild-type templates. DNA oligonucleotides of sequence 5′-TAATACGACTCACTATACCC-3′ and 5′-CCCACCCAAAAAAAAAAAAAAAAAGGGGGGTATAGTGAGTCGTATTA-3′ formed the iC$_6$ template, 5′-TAATACGACTCACTATAAAA-3′ and 5′-CCCAC CCAAAAAAAAAAAAAAAAATTTTTTTTATAGTGAGTCGTATTA-3′ formed the iA$_6$ template and 5′-TAATACGACTCACTATAGGG-3′ and 5′-CCCACCCAA AAAAAAAAAAAAAATCTCCCTATAGTGAGTCGTATTA-3′ formed the wild-type template. Each 200 μl transcription reaction was treated with 2 μl calf intestinal phosphatase for 1 h at 37 °C, extracted with phenol-chloroform twice to remove enzymes and precipitated with ethanol. Pellets were resuspended in 1× DNase Turbo buffer with 5 μl DNase in a total volume of 100 μl, incubated for 2 h at 37 °C, extracted with phenol-chloroform twice and precipitated with ethanol. Purified transcript (5 μl) was mixed with 1 μl T4 polynucleotide kinase (in 1× polynucleotide kinase buffer) and incubated at 37 °C for 1 h. Treated RNA (1 μl) was ligated to 30 ng RNA adaptor of sequence 5′-GCUGAUGGC GAUGAAUGAACACUGCGUUUGCUGGCUUUGAUGAAA-3′ with T4 RNA ligase in 1× RNA ligase buffer at 37 °C for 1 h. Ligated RNA (1 μl) was reverse transcribed by mixing with 1 mM deoxynucleotide triphosphates, and a complementary DNA primer of sequence 5′-CCCCCAAACCCCCAAAAAAAAAAAC CCACCCAAAAAAAAAAAAAA-3′ at a final concentration of 5 μM at 65 °C for 15 min followed by 10 U μl^{-1} SuperScript III reverse transcriptase in 1× reverse transcriptase buffer, 5 mM MgCl$_2$ and 10 mM dithiothreitol at 50 °C for 1 h. Enzymes were denatured at 85 °C for 5 min followed by DNA amplification with sequential PCR reactions using primers 5′-CGATCCGAACGCAGCATTTAC GCTGATGGCGATGAATGAACACTG-3′ and 5′-CCCCCAAACCCCCAAAA AAAAAAACCCACCCAAAAAAAAAAAAAA-3′, digestion with MlyI and HinfI in NEBuffer 4 to cleave sequences containing the promoter, and PCR amplification with 5′-GCTAGTTATTGCTCAGCGGAATAACGATCCGAACGCAGCATTT AC-3′ and 5′-GCTAGTTATTGCTCAGCGGAAAAAAAAAAAAAAACCCCCAA ACCCCC-3′. PCR products were cloned into the backbone of plasmid AP-T7 P amplified with primers 5′-CGGATCGTTATTCCGCTGAGCAATAACTAGC AGAGCAAAAGGCCAGC-3′ and 5′-GGGGGTTTGGGGGTTTTTTTTTTTTTT CCGGCCTTGTCGGCCTTAC-3′ by isothermal assembly cloning. Individual colonies were picked and sequenced using a primer of sequence 5′-CAGGAA GGCAAAATGCCG-3′.

30. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods* **6,** 343–345 (2009).
31. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* **97,** 6640–6645 (2000).
32. Ichetovkin, I. E., Abramochkin, G. & Shrader, T. E. Substrate recognition by the leucyl/phenylalanyl-tRNA-protein transferase. Conservation within the enzyme family and localization to the trypsin-resistant domain. *J Biol Chem* **272,** 33009–33014 (1997).

CORRECTIONS & AMENDMENTS

Water and its influence on the lithosphere–asthenosphere boundary

David H. Green, William O. Hibberson, István Kovács & Anja Rosenthal

This Letter redetermined the solidus of the Earth's mantle in the presence of small amounts of water, and explored the stability of hydrous minerals at the solidus and water solubility in nominally anhydrous minerals (NAMs). In Supplementary Fig. 1, we now illustrate a disagreement between our experimental study and an example of a modelling approach[1]. The modelling approach[1] relies on three factors that must be experimentally determined: (1) the vapour-saturated solidus of lherzolite + H_2O as a function of P,T; (2) the partitioning of water between residual phases (hydrous and nominally anhydrous) and melt; and (3) the liquidus depression of olivine-rich melts as a function of increasing water content. Our experimental study provides points (1) and (2). The water partitioning ($D_{\mathrm{minerals/melt}}$ ~0.0007) obtained in our near-solidus experiments at 2.5 GPa, 1,050 °C, is an order of magnitude lower than that used in the modelling approach in ref. 1. The higher values used in the modelling approach[1] are based on mineral/melt partitioning measured at higher temperatures and much lower water contents in the melt phase. There are unresolved questions on the different techniques used for water analysis in NAMs (bulk water versus structurally bound water) and on the effects on $D_{\mathrm{minerals/melt}}$ of temperature, and of mineral and melt compositions. Because of the importance of mineral/melt water partitioning to the modelling approach[1], it would have been helpful to have included some examples in our Letter of the detailed phase compositions and measured water contents in NAMs. These data establish that the water content in residual mantle at the water-saturated solidus is ~180 p.p.m. In Supplementary Tables 1 and 2 and Supplementary Fig. 2 we have now added examples from our full data set, which will be published in the petrological literature. We thank E. H. Hauri for bringing this to our attention.

1. Katz, R. F., Spiegelman, M. & Langmuir, C. H. A new parameterization of hydrous mantle melting. *Geochem. Geophys. Geosyst.* **4,** 1073 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

CORRECTIONS & AMENDMENTS

Extended megadroughts in the southwestern United States during Pleistocene interglacials

Peter J. Fawcett[1], Josef P. Werne[2,4,5], R. Scott Anderson[6,7], Jeffrey M. Heikoop[8], Erik T. Brown[3], Melissa A. Berke[3], Susan J. Smith[7], Fraser Goff[1], Linda Donohoo-Hurley[1], Luz M. Cisneros-Dozal[8], Stefan Schouten[9], Jaap S. Sinninghe Damsté[9], Yongsong Huang[10], Jaime Toney[10], Julianna Fessenden[8], Giday WoldeGabriel[8], Viorel Atudorei[1], John W. Geissman[1] & Craig D. Allen[11]

[1]Department of Earth & Planetary Sciences, University of New Mexico, Albuquerque, New Mexico 87131, USA. [2]Large Lakes Observatory and Department of Chemistry and Biochemistry, University of Minnesota Duluth, Duluth, Minnesota 55812, USA. [3]Large Lakes Observatory and Department of Geological Sciences, University of Minnesota Duluth, Duluth, Minnesota 55812, USA. [4]Centre for Water Research, University of Western Australia, Crawley, Western Australia 6009, Australia. [5]WA-Organic and Isotope Geochemistry Centre, Curtin University of Technology, Bentley, Western Australia 6845, Australia. [6]School of Earth Sciences and Environmental Sustainability, Northern Arizona University, Flagstaff, Arizona 86011, USA. [7]Laboratory of Paleoecology, Bilby Research Center, Northern Arizona University, Flagstaff, Arizona 86011, USA. [8]Earth and Environmental Sciences Division, EES-14Los Alamos National Laboratory, Los Alamos, NewMexico 87545, USA. [9]NIOZ Royal Netherlands Institute for Sea Research, Department of Marine Organic BiogeochemistryPO Box 59, 1790 AB Den Burg, Netherlands. [10]Department of Geological Sciences, Brown University, Providence, Rhode Island 02912, USA. [11]USGS Fort Collins Science Center, Jemez Mountains Field Station, Los Alamos, New Mexico 87544, USA.
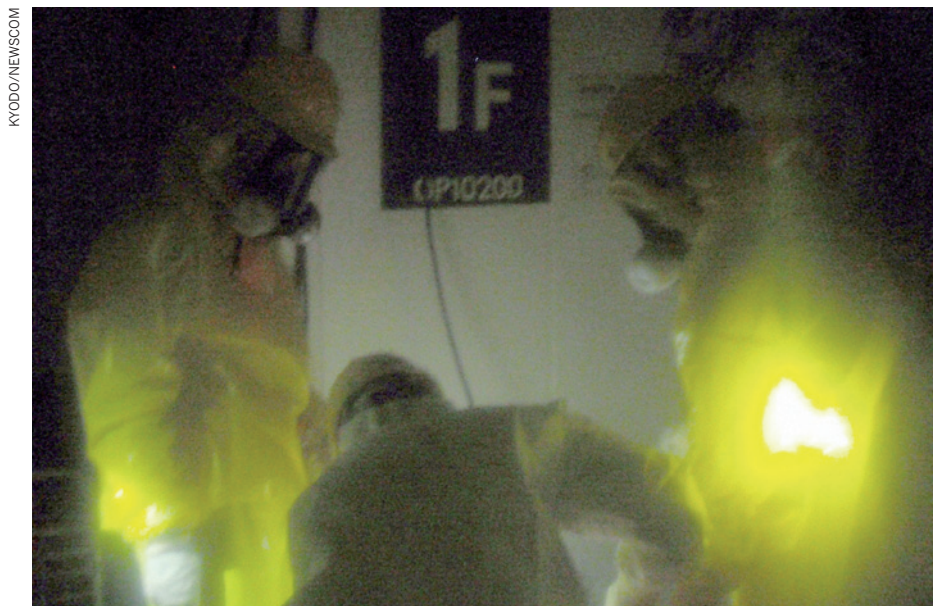
In this Letter, the affiliations for authors Jaime Toney, Julianna Fessenden and Giday WoldeGabriel were listed incorrectly. The correct affiliations for all authors are as shown here.

CAREERS

Japan's nuclear crisis is expected to boost demand for nuclear specialists globally.

NUCLEAR ENERGY

Defying disaster

The worst nuclear disaster since Chernobyl is unlikely to slow job growth in the industry.

BY QUIRIN SCHIERMEIER

The catastrophe at Japan's Fukushima Daiichi nuclear power plant is likely to have long-term ramifications for the nuclear industry. It is not, however, expected to shrink the number of jobs in the field. In fact, in some niches, such as safety and quality control, the number may even rise.

Despite the crisis, many countries still consider nuclear power an important part of their energy future. Nuclear specialists are therefore likely to remain in high demand. "The tragic events in Japan made everyone take pause, but they will not stop the recent nuclear renaissance," says Elizabeth McAndrews-Benavides, programme manager for workforce development at the Nuclear Energy Institute in Washington DC.

In response to the events in Japan, governments in the United States, Asia and the European Union have ordered additional safety checks that could lead to modifications in plant design and operation. These changes are likely to increase, not decrease, the demand for nuclear professionals, says McAndrews-Benavides. Moreover, says Weston Stacey, a nuclear engineer at the Georgia Institute of Technology in Atlanta, the industry will soon have to replace the thousands of engineers and technicians who are nearing retirement (see *Nature* **459,** 124–125; 2009). In the United States alone, some 38% of the nation's nuclear workforce will be eligible for retirement by 2015, he says.

Many US universities already offer undergraduate to doctoral programmes in nuclear engineering that provide students with solid knowledge and skills in reactor physics, engineering and materials science. Heavily recruited by local utilities and national labs, graduates of these programmes have a promising future, says Stacey.

In Japan, where funding for (and training in) nuclear engineering is detached from normal university funding, scientists expect that collaboration between university researchers and industry specialists will expand in response to the country's triple disaster of earthquake, tsunami and near nuclear meltdown. "This tragedy must make us rethink how we deal with, and respond to, nuclear risks, says Shoji Nagamiya, director of the Japan Proton Accelerator Research Complex (J-PARC). In addition to reactor design and safety, Nagamiya predicts that scientists will be needed to work on health issues such as the effects of nuclear radiation on foods, animals and humans. In some countries, the most challenging obstacle will be to get science students enthusiastic about nuclear energy and technology. In Germany, where the public is notoriously hostile towards nuclear energy, the government ordered that the seven oldest plants be switched off temporarily, and put a three-month moratorium on the previously planned extension of its 17 nuclear reactors.

But even though it is phasing out nuclear energy, Germany must train experts who can cogently and clearly present data on technical nuclear issues to regulators to help them incorporate the information into policy, says Joachim Knebel, chief science officer at the Karlsruhe Institute of Technology in Germany, who has helped set up six working groups to assess the accidents at the Fukushima plant and outline future research needs. "Well-trained nuclear engineers and safety experts will always find jobs, whether in this country or abroad," he says.

The Organisation for Economic Co-operation and Development (OECD) estimates that the 14% of electricity currently produced by nuclear means worldwide will rise by 10–20% by 2050. Sixty-five nuclear plants are currently under construction, 27 in China alone. The United States is building one new reactor, and up to eight are scheduled to go online by 2020. Whether safety reviews ordered in response to the Fukushima events will slow the growth of the global nuclear fleet is not yet clear.

The briskly rising global demand for energy will require all available technologies, including nuclear, says Thierry Dujardin, deputy director for science and development at the Nuclear Energy Agency in Paris. "So whatever nuclear policies some countries may opt for," he says, "the world will need a qualified nuclear workforce for decades to come." ∎

FUKUSHIMA CRISIS

➔ WWW.NATURE. COM/JAPANQUAKE

PHYSICS
UK launches fellowships

The UK Science and Technology Facilities Council (STFC), a government funding agency, is creating a five-year physics and astronomy fellowship programme. Under the Ernest Rutherford scheme, 12 fellows will be appointed each year until there are 60, a number that will then be maintained. The pay rate is undisclosed, but fellowships are open to residents and non-residents and can be taken at a UK university of the fellow's choice. Fellows can apply for extra funding from an STFC pot that will total £3 million (US$4.9 million) by 2016. Applicants must be early-career astronomers or particle or nuclear physicists who can head a research group and have a track record of leadership. Applications will open by July, and decisions will be made in early 2012.

RESEARCH AND DEVELOPMENT
Asian spending grows

China, Turkey, Iran and other Asian countries are spending more on research, says a report, making creation of science jobs likely. *Knowledge, Networks and Nations*, released on 28 March by the Royal Society in London, finds that China expects to spend 2.5% of its gross domestic product (GDP) on research by 2020, and India hopes to reach that goal by 2022. South Korea has pledged that research spending will reach 5% of GDP by 2012. Turkey increased its expenditure six-fold between 1995 and 2007, and Iran is forging industry–academia links in nanotechnology and biotechnology. The report recommends easing cross-border movement to help scientists fill the research posts that will arise as a result.

COMPUTER SCIENCE
Postdoc prevalence rises

A growing proportion of US computer-science PhD holders are pursuing postdoctoral research or industry rather than tenured academia, says a report by the Computing Research Association (CRA) in Washington DC. *The Role of PostDocs in Computer Science*, out earlier this year, says that the number of new doctorates who were hired for tenure-track posts fell by one-third from 2004 to 2009, yet that hired by industry tripled. Postdoc appointments grew three-fold between 2001 and 2009. The report asks whether the field will come to resemble the life sciences, with researchers doing several postdocs before securing a permanent academic job.

TURNING POINT
Ian Chapman

Ian Chapman, a nuclear physicist at the Culham Centre for Fusion Energy in Abingdon, UK, won the Cavendish Medal at the Science, Engineering and Technology Student of the Year Awards in London in March, after proving his ability to convey the significance of his research to Parliament.

Describe the work that won you the award.
To get energy from nuclear fusion, you must heat the hydrogen fuel to 150 million °C — ten times hotter than the centre of the Sun — so that the isotopes can fuse. But when you heat an ionized gas, or plasma, that much, it can become unstable. My job is to understand when and how the instabilities appear, and how to get rid of them. One example is a periodic collapse in the temperature and density of the plasma. The administrators of ITER, an international project to build an experimental fusion reactor, asked the community to investigate this instability. I chaired a collaboration of 16 labs to model the instability and demonstrate that we understand what is required to control it.

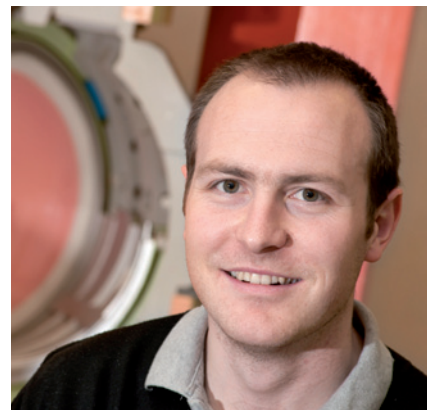**Are communication skills important for fusion researchers?**
Yes. Communicating what we do is a problem for physicists in general. It's absolutely important. Awards like this help to raise public understanding of fusion, which can only be good for future funding. And, when working on large international teams, I've found that I'm good not only at generating enthusiasm, but also at helping partners to tackle problems together.

What's the status of ITER?
It is being built in the south of France, and will prove once and for all whether fusion is viable — specifically, whether we can get more power out than is put in. The first plasma in which fusion can occur should be produced in November 2019. In terms of the size and number of collaborations, ITER is similar to the Large Hadron Collider (the world's largest high-energy particle accelerator, located at the CERN particle-physics lab near Geneva, Switzerland), but some things have yet to be worked out — for example, whether researchers will remain at their home institutions or become ITER employees.

Do you have to be involved now to have a role at ITER in the future?
No. All the partners are collaborating to make ITER work so that there will continue to be opportunities for research and jobs. Those with the best ideas will get the jobs.

What challenges does ITER face?
ITER is set up so that all knowledge about how to design, assemble and run the experiment is shared among the collaborators. For example, the biggest component of ITER is the field magnets. Rather than having one partner provide those, it was decided that everyone would be involved in manufacturing them so that everyone would have the knowledge to build their own fusion power plants in future. That is a sensible long-term strategy, but it adds a lot of negotiations and complexity to the project.

What have you done to set yourself apart from other young researchers?
The vast majority of machines to heat plasma use neutral-beam injection, which spins the plasma at hundreds of kilometres per second. A mentor pointed out that nobody was looking into the effects of that on the plasma. We worked out how to model the rotational dynamics, which gave us a unique skill.

Are younger scientists flocking to fusion energy now that ITER is a reality?
Yes, the calibre of PhD students and postdocs interested in fusion has increased dramatically in the past few years. We are getting more publicity now that ITER is really going ahead, and is no longer simply a paper exercise. This is the most exciting time to work in fusion, because the next decade of work will make or break the field — and define whether fusion can work for humankind. ∎

INTERVIEW BY VIRGINIA GEWIN

UKAEA

A GOOD TIME

A real body blow.

BY SHELLY LI

Whenever he got drunk, my old man would tell me: "If you're looking to get married and have kids and a white picket fence, get a girlfriend. If you got a girlfriend when all you're looking for is sex … you're overpaying." Sometimes I wonder if he was the same guy before Ma left him.

Driving down the street, I can feel the car engine humming in my fingertips as I grip the wheel. During the day, I follow the rules and let the car drive itself. It speeds up the traffic, I know. But at night, I love sitting in the driver's seat, tilting my wrist left and right, and watching the car heed my will. This baby is one of the only things I've got to show for five years and counting in investment banking.

Stoplight colours blur together at the speed that I'm going. Everything in my peripherals has long folded into a grey vortex.
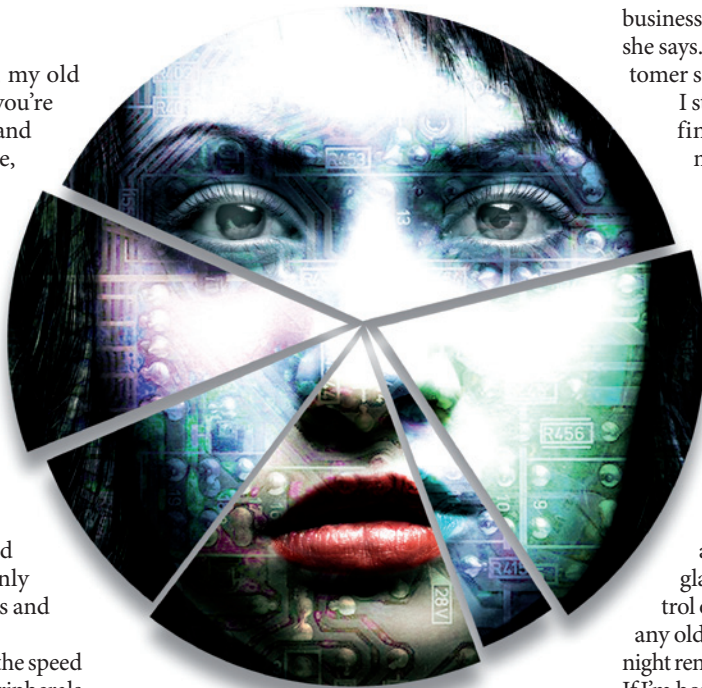
Thirty miles later, I find myself downtown, and I ease onto the brake. Cops don't usually hang around here, but I slow down anyway to enjoy the moment.

The lights, the scarce number that there are on this street, solidify to a standstill. All along the walls, young women line up like prisoners waiting to be decimated. They stand casually, though, a nonchalant shrug in their shoulders, a carefree glint in their eyes. You won't find a girl more beautiful than one on this street. You also won't find a real one — real girl, I mean.

Prostitution only became legal after the robotics companies took over the business. It was just more tax money to collect, so the mayor didn't seem to mind. Who knows? Maybe the potholes in my neighbourhood were finally fixed last year at the expense of some socially stunted invalid who scheduled daily appointments of sexual therapy?

I pull up by the kerb and roll down my window. Immediately, a pretty little thing steps forward and fills the frame. "Hey there," she says, the fine tips of her dark hair brushing my left arm. She looks over the inside of my car first, checking to see how much I'll be capable of paying. Quality and price are adjusted per client. "Whatcha lookin' for, handsome?"

I open my mouth to tell her that I'm looking for an escape, but all that comes out are the cliché words I'm used to saying. "A good time."

The woman smiles before slowly raising her head to make eye contact. Her blue-grey eyes are arresting, and looking at them is like viewing the ocean from space, watching drifting silver clouds blanket a massive body of glittering blue. The company, whichever one she belongs to, must have spent extra time producing this model.

Finally she says: "Sorry, not selling that today." And with that, she backs away from my car and turns to leave.

"Hey, wait," I call out. For a moment my mind freezes, and I have no idea what to say as she turns around to face me again. Did that just happen? "You're not allowed to turn me down."

The woman blinks. "For the next month, I will be. Companies are ordered to regenerate their models."

Frustration courses through my blood, pent-up anger rising from God knows where.

"Regenerate what? This is not right. I want to speak with your … manager — I mean supervisor — I mean, pimp. Whatever you call him."

At this the woman lets out a warm and husky laugh. She walks forward and hands me a business card. "Call the number on the card," she says. "Or you can send a message to customer service."

I stare blankly at the card between my fingers, shocked and confused and maybe even a little hurt. Sitting in my 760 BMW, window rolled down, a piercing autumn wind seeping through my sweater, Pa's voice pops into my head. I can just imagine him muttering quietly to me, the scent of Myer's Rum lacquered to his breath: "Son, you know you've reached a new low when a whore tells you no."

After a long pause, the woman starts to walk away again.

My stomach does a flip, and I call after her. "You selling anything at all?" The desperation in my voice is glaring, one that I cannot seem to control or understand. If I want, I can call up any old college girlfriend and set up a late-night rendezvous. This is about more than sex. If I'm honest, it's always about more than sex.

The woman sighs, her breath curling up as a silver wisp in the night air. She puts her hands on her hips and turns around, looking annoyed now. It suddenly occurs to me that I do not know whether she can feel or not. Robots can surely reason, so why shouldn't they be able to feel? I wonder why I never bothered to wonder this before.

"What do you do for a living?" she asks me.

I frown at the question. "I'm an investment banker. Why?"

"Oh, fantastic. This'll be easy to explain." She takes a few steps closer to my car, but stays a casual five feet away. "Supply and demand. In order to keep the demand high, the government must limit supply. You and I, we're all just part of the business, sugar."

This time, when she walks away, I make no attempt to stop her. Soon she disappears around a corner, becoming a shadow against an adjacent wall, and then a phantom.

I roll my window up and lean back against my car seat. A minute or so passes, and I begin to chuckle at my foolishness. Hitting the side of my steering wheel, I drive off. The lights all blur together again.

The microeconomics of prostitution. Now there's a class I missed in graduate school. ∎

Shelly Li *has published more than 20 short stories. Her first novel,* The Royal Hunter, *is forthcoming from Philomel Books. For more information, visit: www.shelly-li.com.*

LETTER

Spin crossover and iron–rich silicate melt in the Earth's deep mantle

Ryuichi Nomura[1,2], Haruka Ozawa[1,3], Shigehiko Tateno[1], Kei Hirose[1,3], John Hernlund[4], Shunsuke Muto[5], Hirofumi Ishii[6] & Nozomu Hiraoka[6]

A melt has greater volume than a silicate solid of the same composition. But this difference diminishes at high pressure, and the possibility that a melt sufficiently enriched in the heavy element iron might then become more dense than solids at the pressures in the interior of the Earth (and other terrestrial bodies) has long been a source of considerable speculation[1,2]. The occurrence of such dense silicate melts in the Earth's lowermost mantle would carry important consequences for its physical and chemical evolution and could provide a unifying model for explaining a variety of observed features in the core–mantle boundary region[3]. Recent theoretical calculations[4] combined with estimates of iron partitioning between (Mg,Fe)SiO$_3$ perovskite and melt at shallower mantle conditions[5–7] suggest that melt is more dense than solids at pressures in the Earth's deepest mantle, consistent with analysis of shockwave experiments[8]. Here we extend measurements of iron partitioning over the entire mantle pressure range, and find a precipitous change at pressures greater than ~76 GPa, resulting in strong iron enrichment in melts. Additional X-ray emission spectroscopy measurements on (Mg$_{0.95}$Fe$_{0.05}$)SiO$_3$ glass indicate a spin collapse around 70 GPa, suggesting that the observed change in iron partitioning could be explained by a spin crossover of iron (from high-spin to low-spin) in silicate melt. These results imply that (Mg,Fe)SiO$_3$ liquid becomes more dense than coexisting solid at ~1,800 km depth in the lower mantle. Soon after the Earth's formation, the heat dissipated by accretion and internal differentiation could have produced a dense melt layer up to ~1,000 km in thickness underneath the solid mantle. We also infer that (Mg,Fe)SiO$_3$ perovskite is on the liquidus at deep mantle conditions, and predict that fractional crystallization of dense magma would have evolved towards an iron-rich and silicon-poor composition, consistent with seismic inferences of structures in the core–mantle boundary region.

Our melting experiments were performed on samples with bulk composition (Mg$_{0.89}$Fe$_{0.11}$)$_2$SiO$_4$ at pressures from 20 to 159 GPa in a laser-heated diamond-anvil cell (DAC; Supplementary Table 1). The heating duration was short in order to avoid anomalous thermal diffusion (Supplementary Information), but this prevented us from measuring the melting temperature. Nevertheless, the upper and lower bounds of the temperature in our experiment are given by the liquidus temperature of Mg$_2$SiO$_4$ and the solidus temperature of natural peridotite, respectively (see Methods and Supplementary Fig. 1). Samples were recovered from the DAC and examined with a high-resolution field-emission-type electron probe micro-analyser (FE-EPMA). Recovered specimens exhibited a concentric texture that reflected the temperature distribution during heating (Fig. 1), which is similar to that observed in conventional multi-anvil experiments[5–7]. We consistently found a pocket at the hottest part of the sample that possessed non-stoichiometric composition, which we interpret as quenched partial melt. The (Mg+Fe)/Si molar ratio of this quenched melt increased with pressure, from 1.50 at 36 GPa to 2.56 at 159 GPa (Supplementary Fig. 2). The melt pocket was surrounded by a single-phase solid layer (ferropericlase or perovskite, depending on pressure), which we interpret to be
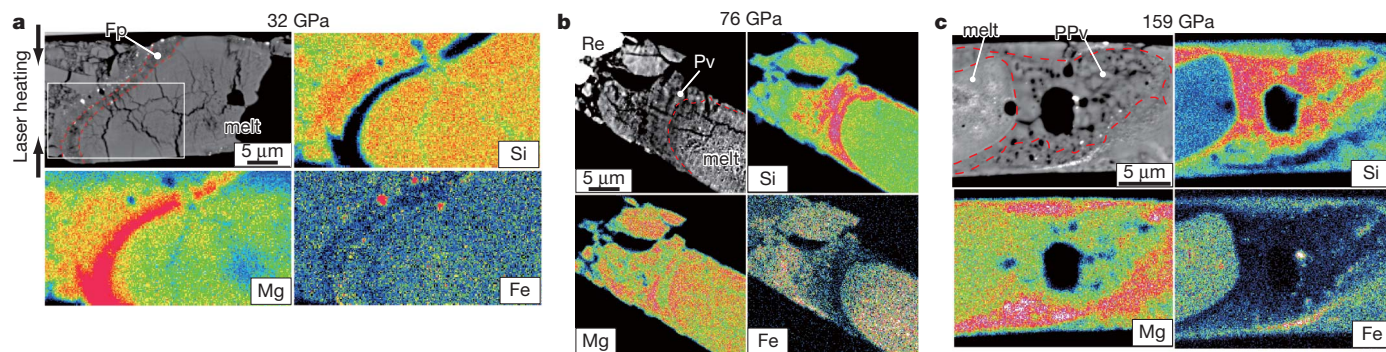

Figure 1 | Backscattered electron images and X-ray maps for Si, Mg and Fe for samples recovered from high-pressure melting experiments. **a**, At 32 GPa, when ferropericlase (Fp) is the liquidus phase; **b**, at 76 GPa, when perovskite (Pv) is the crystalline phase in contact with the quenched melt pocket; and **c**, at 159 GPa in the stability field of post-perovskite (PPv). Quenched melt was found at the centre of the sample, where the temperature was highest. Metallic iron was observed at the edge of the laser-heated area in all samples, where a strong temperature gradient existed[28]. It was also found in the melt pocket, but only above 36 GPa where the liquidus phase was perovskite or post-perovskite. Arrows in **a** represent the directions of the laser beams for heating. See Supplementary Information for the valence state of iron in the partial melt.

[1]Department of Earth and Planetary Sciences, Tokyo Institute of Technology, Meguro, Tokyo 152-8551, Japan. [2]Department of Earth and Planetary Sciences, University of Tokyo, Bunkyo, Tokyo 113-0033, Japan. [3]Institute for Research on Earth Evolution, Japan Agency for Marine-Earth Science and Technology, Yokosuka, Kanagawa 237-0061, Japan. [4]Department of Earth and Planetary Science, University of California, Berkeley, California 94720, USA. [5]Department of Materials, Physics and Energy Engineering, Nagoya University, Chikusa, Nagoya 464-8603, Japan. [6]National Synchrotron Radiation Research Center, Hsinchu 30076, Taiwan.

the phase that crystallizes first as temperature decreases (the liquidus phase)[5,7].

In this study, the liquidus phase was ferropericlase from 20 to 36 GPa and was subsequently replaced by perovskite at higher pressures (both were in contact with the melt pool at 36 GPa) (Fig. 1). Considering the difference in Mg/Si ratio of the samples[7], this is consistent with observations[5] of material with peridotite composition where the liquidus phase changes from ferropericlase to perovskite above 31 GPa. Although phase identification was not made, post-perovskite should have been formed in our experiments performed at 143 and 159 GPa. The change in liquidus phase from ferropericlase to perovskite at 36 GPa indicates that the eutectic melt composition becomes more Mg-rich at higher pressures. This is in agreement with the increase in (Mg+Fe)/Si molar ratio of partial melt with increasing pressure (Supplementary Fig. 2), although the partial melt composition also depends on the degree of melting.

The Fe-Mg distribution coefficient between perovskite/post-perovskite and melt, $K_D = ([Fe^{Pv}]/[Mg^{Pv}])/([Fe^{melt}]/[Mg^{melt}])$, was determined in the pressure range from 36 to 159 GPa (Fig. 2a), where perovskite/post-perovskite was th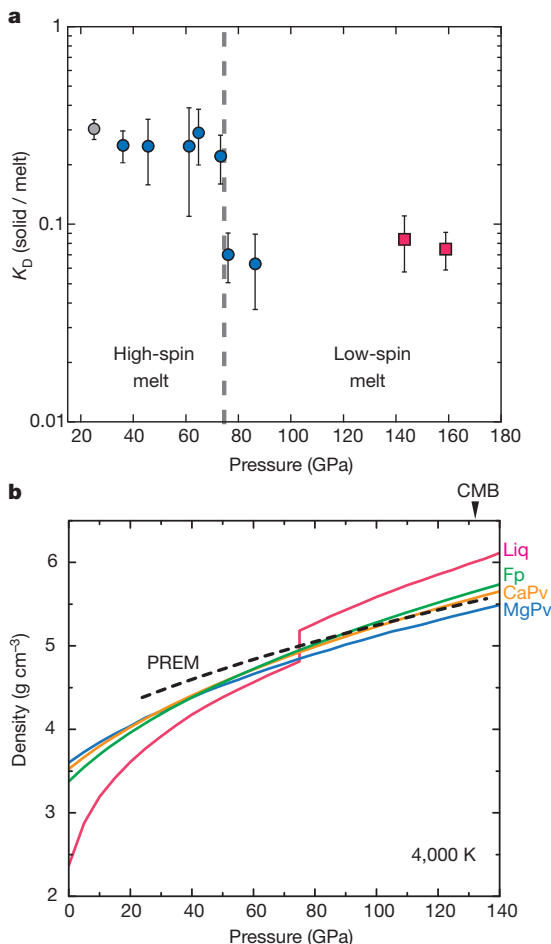e liquidus phase. Although the quenched melt pocket contained multiple valence states of iron (Fig. 1b, c), we consider that all Fe was present as Fe^{2+} at high temperature (see Supplementary Information). The obtained K_D values were approximately constant (at 0.22–0.29) below 73 GPa. These values are somewhat lower than previous determinations (~0.4) in Al-bearing peridotite bulk compositions[5–7] but are in good agreement with $K_D = 0.304 \pm 0.035$ at 25 GPa in Al-free peridotite material[6]. The high K_D in Al-bearing systems should be due to the high Fe^{3+} content of perovskite (see, for example, ref. 9). On the other hand, the value of K_D suddenly dropped to 0.07 ± 0.02 at 76 GPa (Fig. 2a). It then remained almost constant at 0.06–0.08 up to 159 GPa.

In order to explore the cause of strong Fe-enrichment in partial melt at pressures above 76 GPa, we collected the X-ray emission spectra of $(Mg_{0.95}Fe_{0.05})SiO_3$ glass at 300 K at increasing pressure, from 8 to 85 GPa (Fig. 3). At low pressures, the Fe $K\beta'$ satellite peak was clearly observed at 7,045 eV, showing the presence of high-spin Fe^{2+} in the glass sample. The satellite peak diminished somewhat at 59 GPa, and vanished at 77 GPa. This indicates a spin collapse in ferrous iron. Insofar as the glass is a good analogue for the liquid state, such a high-spin to low-spin crossover of iron may also occur in melt at a similar pressure range, thus providing an explanation for the measured jump in Fe-enrichment in partial melt (see, for example, refs 10, 11). Indeed, the pressure range of the spin crossover observed in our glass matches the pressure where K_D changed dramatically (Fig. 2a). Our melts had higher Mg/Si ratios and FeO contents than $(Mg_{0.95}Fe_{0.05})SiO_3$ glass; however, the calculated mean Mg–O and Si–O coordination numbers for Mg_2SiO_4 liquid[12] are very similar to those for $MgSiO_3$ liquid[13] at lower mantle pressures. Therefore, the spin crossover pressure, which depends on Mg/Fe–O coordination, would not shift significantly following a change in Mg/Si ratio in melt from 1 to 2. Theory[14,15] suggests that the Fe content does not change the pressure range of the spin crossover when the Fe concentration is small (less than ~20%), because Fe–Fe interactions are negligible. Multi-anvil experiments[16] have shown a very sharp change in Fe partitioning between perovskite and ferropericlase at the commencement of the spin crossover, which is compatible with the present observations.

A strong change in iron partitioning suggests that partial melt becomes more dense at depths below 1,800 km. We calculated the

Figure 2 | Change in Fe-Mg distribution coefficient and calculated density profiles. a, $K_D = (Fe^{Pv}/Mg^{Pv})/(Fe^{melt}/Mg^{melt})$ between perovskite (blue circles) or post-perovskite (red squares) and melt; the values drop sharply at pressures above 76 GPa, probably due to the effect of the spin crossover of iron in silicate melt (see Fig. 3). Previous experimental datum obtained at 25 GPa using a multi-anvil apparatus is shown by a grey circle[6]. Error bars were estimated from uncertainties (1σ) in both solid and liquid compositions. **b**, Density of the (Mg,Fe)SiO$_3$ liquid coexisting with $(Mg_{0.92}Fe_{0.08})SiO_3$ perovskite calculated for 4,000 K using the newly obtained Fe-Mg partitioning data. Data for $(Mg_{0.86}Fe_{0.14})O$ ferropericlase[29], Ca-perovskite[30] and PREM[19] are also shown for comparison. Liq, liquid; Fp, ferro-periclase; CaPv, calcium silicate perovskite; MgPv, magnesium silicate perovskite.

Figure 3 | Evolution of X-ray emission spectra of $(Mg_{0.95}Fe_{0.05})SiO_3$ glass with increasing pressure. Measurements were conducted at 300 K. All spectra are normalized to transmitted intensity, and shifted so that the weighted average of main (Kβ) plus satellite (Kβ') emission lines is set to 7,058 eV. The satellite peak decreased slightly at 59 GPa and completely disappeared at 77 GPa, indicating the spin crossover of iron.

density of $(Mg,Fe)SiO_3$ liquid in equilibrium with $(Mg_{0.92}Fe_{0.08})SiO_3$ perovskite—a typical composition for pyrolitic lower mantle[16,17]—at 4,000 K as a function of pressure (Fig. 2b). For simplicity, we used $K_D = 0.25$ below 75 GPa and $K_D = 0.07$ at higher pressures. The molar volume of $MgSiO_3$ liquid was obtained by recent first principles calculations[4], and the effect of iron is assumed to be the same for liquid and solid (perovskite)[4,18]. Whereas $(Mg,Fe)SiO_3$ melt is buoyant compared to any of the typical lower-mantle minerals below 75 GPa, it suddenly becomes more dense at higher pressures. The difference from the preliminary reference Earth model (PREM)[19] reaches 8% at the base of the mantle. Partial melts obtained in the present experiments had a higher Mg/Si ratio than $MgSiO_3$ (Supplementary Fig. 2). According to previous shock-wave compression experiments[8], such high-$(Mg+Fe)/Si$ melt is likely to be denser than the $(Mg,Fe)SiO_3$ melt considered above, suggesting that the density crossover between melt and solid may occur even at depths shallower than 1,800 km. Although details of the formation of melts at the base of the mantle are not yet clear, the present results provide a constraint on the plausible thickness of any stable melt layer of around 1,000 km (Fig. 4).

Labrosse et al.[3] have presented a model in which dense melts produced soon after Earth's formation comprised a potentially large 'basal magma ocean' (BMO), which underwent slow crystallization over billions of years, at a rate governed by relatively sluggish solid state convection in the overlying solid mantle. The present experimental results provide a new physical rationale for the gravitational stability of a BMO up to ~1,000 km thick underneath the solid mantle, and the maximum plausible thickness inferred above is broadly consistent with the BMO hypothesis. For example, a BMO of ~1,000 km thickness would comprise about one-quarter of the mantle's mass; fractional crystallization and sequestering of incompatible heat producing elements (for example, U, Th) in the residual liquid as the layer cools would therefore account for the 'missing' chondritic complement of these species expected to be sequestered in a reservoir inside the mantle[20]. The present results also allow us to make further predictions about the nature of the BMO and its chemical evolution through time.

As noted above, our results demonstrate that $(Mg,Fe)SiO_3$ perovskite is the first phase to crystallize from melts with a wide range of $(Mg+Fe)/Si$ ratios; this is true even for Si-poor melt with $(Mg+Fe)/Si \approx 2.5$ at the conditions of the core–mantle boundary (Supplementary Fig. 2). This strongly suggests that an episode of perovskite crystallization might well characterize the majority of the freezing history of a BMO, with other phases such as $(Mg,Fe)O$ magnesiowüstite only crystallizing relatively late. In addition, perovskite crystals forming in the BMO would have been relatively depleted in iron and floated to the top of magmas at depths greater than 1,800 km (Fig. 4), because of the small K_D value for Fe/Mg partitioning. As a consequence, we can predict that the residual magma would evolve towards a FeO-rich/SiO_2-poor composition (that is, near the composition of wüstite), becoming more dense with time and probably also retaining a variety of incompatible volatile species.

Evolution through fractional crystallization as described above would also have affected the composition of cumulates that formed from the BMO. In particular, cumulates should become more Fe-rich with time, and presumably more dense, as they crystallize from an increasingly Fe-rich magma. These dense cumulates themselves might eventually become stable against complete entrainment by mantle convection currents, and pile up at the base of the mantle to form thermo-chemical piles (Fig. 4). The accumulation of dense solid material, which is around 2–3% denser than average mantle and which comprises ~2% of the total mantle volume, can explain the presence of two large, low-shear-wave velocity provinces (LLSVPs) at the base of the mantle beneath the Pacific and Africa[21]; and the expected Fe-enriched composition of the dense material is consistent with the magnitude of these anomalies[22].

Larger degrees of crystallization in a BMO could also have left patches of dense residual mush in thin layers above the core–mantle
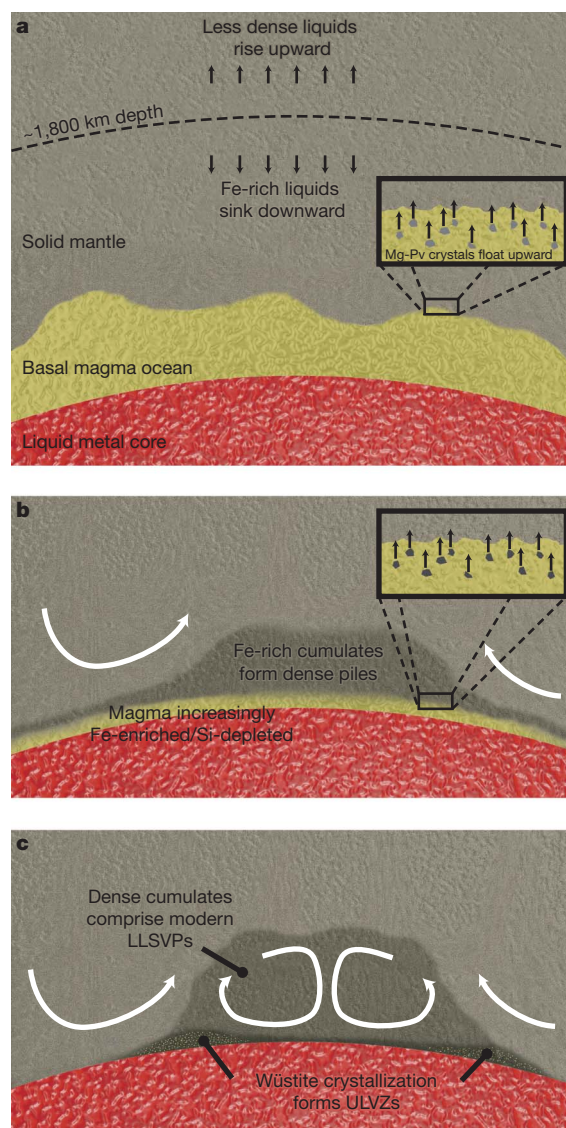

Figure 4 | Evolution and crystallization of dense melts in the deep mantle. **a**, During Earth's early history, any melts that form below ~1,800 km depth sink and accumulate at the base of the mantle, while any crystals that form owing to cooling of this dense magma will rise upward into the solid mantle. **b**, Fe-poor perovskite crystallization leaves a residual liquid enriched in FeO and depleted in SiO_2, and crystals forming from this evolved liquid may become dense enough to form thermo-chemical piles at the base of the solid mantle. **c**, The final stage of crystallization involves a composition close to wüstite, leaving behind a very dense thin layer that is consistent with the seismic properties inferred inside ULVZs. White arrows indicate schematic flow patterns in the convecting solid mantle.

boundary, which would explain the presence and seismic velocities of ultralow-velocity zones (ULVZs)[23]. Such material would be maintained at or near the solidus over geological timescales because the residual liquid would sequester incompatible species that in turn depress the melting temperature. Alternatively, late crystallization of wüstite-rich cumulates from the Fe-rich/Si-poor melt could have left behind a dense solid layer with seismic properties that may also be consistent with seismic inferences for ULVZ, even in the absence of melt[24]. The BMO model, together with the evolution of composition that we predict from our experiments, is compatible with both possibilities. Another possibility that might help to explain why ULVZs in

different regions do not always exhibit the same seismic signatures[25] is that inside some ULVZs the interstitial melt has drained out, leaving a largely melt-free solid rock strongly enriched in wüstite. In other ULVZs, owing to different dynamical conditions in the surrounding mantle, continual stirring of the mushy residue by viscous coupling to flows in the overlying mantle might have prevented compaction of the rock and removal of melt[26]. The primary difference we would expect to observe between a melt-free, wüstite-rich ULVZ and a persistently mushy ULVZ would be the strength of the reduction in shear wave velocity, with a greater reduction inside mushy ULVZs[27].

METHODS SUMMARY

We loaded a platelet of $(Mg_{0.89}Fe_{0.11})_2SiO_4$ olivine starting material into the DAC together with argon pressure medium. After compression, the sample was heated by laser to induce melting. Subsequently it was recovered from the DAC and polished parallel to the compression axis. We obtained chemical compositions of the quenched liquid pool and the neighbouring solid phase (liquidus phase) from FE-EPMA analyses of these sections. In addition, several samples were further processed to a thin film and examined under a transmission electron microscope (TEM). The $Fe^{3+}/(Fe^{2+} + Fe^{3+})$ ratio of the quenched melt pocket was determined by electron energy-loss near-edge structure (ELNES) measurements. We also carried out the X-ray emission spectroscopy (XES) study of $(Mg_{0.95}Fe_{0.05})SiO_3$ glass in the DAC at high pressure and room temperature. The high-resolution spectra of the Fe Kβ line were collected through the cubic boron nitride + beryllium composite gasket, and indicate the change in the spin state of iron with increasing pressure.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Stolper, E., Walker, D., Hager, B. H. & Hays, J. F. Melt segregation from partially molten source regions: the importance of melt density and source region size. *J. Geophys. Res.* **86,** 6261–6271 (1981).
2. Agee, C. B. Crystal-liquid density inversions in terrestrial and lunar magmas. *Phys. Earth Planet. Inter.* **107,** 63–74 (1998).
3. Labrosse, S., Hernlund, J. W. & Coltice, N. A crystallizing dense magma ocean at the base of the Earth's mantle. *Nature* **450,** 866–869 (2007).
4. Stixrude, L. et al. Thermodynamics of silicate liquids in the deep Earth. *Earth Planet. Sci. Lett.* **278,** 226–232 (2009).
5. Ito, E., Kubo, A., Katsura, T. & Walter, M. J. Melting experiments of mantle materials under lower mantle conditions with implications for magma ocean differentiation. *Phys. Earth Planet. Inter.* **143–144,** 397–406 (2004).
6. Corgne, A. et al. Silicate perovskite-melt partitioning of trace elements and geochemical signature of a deep perovskitic reservoir. *Geochim. Cosmochim. Acta* **69,** 485–496 (2005).
7. Liebske, C. et al. Compositional effects on element partitioning between Mg-silicate perovskite and silicate melts. *Contrib. Mineral. Petrol.* **149,** 113–128 (2005).
8. Mosenfelder, J. L., Asimov, P. D. & Ahrens, T. J. Thermodynamic properties of Mg_2SiO_4 liquid at ultra-high pressures from shock measurements to 200 GPa on forsterite and wadsleyite. *J. Geophys. Res.* **112,** B06208, doi:10.1029/2006JB004364 (2007).
9. McCammon, C. Perovskite as a possible sink for ferric iron in the lower mantle. *Nature* **387,** 694–696 (1997).
10. Badro, J. et al. Iron partitioning in Earth's mantle: toward a deep lower mantle discontinuity. *Science* **300,** 789–791 (2003).
11. Auzende, A.-L. et al. Element partitioning between magnesium silicate perovskite and ferropericlase: new insights into bulk lower-mantle geochemistry. *Earth Planet. Sci. Lett.* **269,** 164–174 (2008).
12. de Koker, N. P., Stixrude, L. & Karki, B. B. Thermodynamics, structure, dynamics, and freezing of Mg_2SiO_4 liquid at high pressure. *Geochim. Cosmochim. Acta* **72,** 1427–1441 (2008).
13. Karki, B. B. First-principles molecular dynamics simulations of silicate melts: structural and dynamical properties. *Rev. Mineral. Geochem.* **71,** 355–389 (2010).
14. Tsuchiya, T., Wentzcovitch, R. M., da Silva, C. R. S. & de Gironcoli, S. Spin transition in magnesiowüstite in Earth's lower mantle. *Phys. Rev. Lett.* **96,** 198501 (2006).
15. Bengtson, A., Persson, K. & Morgan, D. Ab initio study of the composition dependence of the pressure-induced spin crossover in perovskite $(Mg_{1-x}Fe_x)SiO_3$. *Earth Planet. Sci. Lett.* **265,** 535–545 (2008).
16. Irifune, T. et al. Iron partitioning and density change of pyrolite in Earth's lower mantle. *Science* **327,** 193–195 (2010).
17. Hirose, K. Phase transition in pyrolitic mantle around 670-km depth: implications for upwelling of plumes from the lower mantle. *J. Geophys. Res.* **107,** 2078, doi:10.1029/2001JB000597 (2002).
18. Lundin, S. et al. Effect of Fe on the equation of state of mantle silicate perovskite over 1 Mbar. *Phys. Earth Planet. Inter.* **168,** 97–102 (2008).
19. Dziewonski, A. M. & Anderson, D. L. Preliminary reference Earth model. *Phys. Earth Planet. Inter.* **25,** 297–356 (1981).
20. Hofmann, A. W. Mantle geochemistry: the message from oceanic volcanism. *Nature* **385,** 219–229 (1997).
21. Hernlund, J. W. & Houser, C. On the distribution of seismic velocities in Earth's deep mantle. *Earth Planet. Sci. Lett.* **265,** 423–437 (2008).
22. Cobden, L. et al. Thermochemical interpretation of 1-D seismic data for the lower mantle: the significance of nonadiabatic thermal gradients and compositional heterogeneity. *J. Geophys. Res.* **114,** B11309, doi:10.1029/2008JB006262 (2009).
23. Williams, Q. & Garnero, E. J. Seismic evidence for partial melt at the base of Earth's mantle. *Science* **273,** 1528–1530 (1996).
24. Wicks, J. K., Jackson, J. M. & Sturhahn, W. Very low sound velocities in iron-rich (Mg,Fe)O: implications for the core-mantle boundary region. *Geophys. Res. Lett.* **37,** L15304, doi:10.1029/2010GL043689 (2010).
25. Thorne, M. S. & Garnero, E. J. Inferences on ultralow-velocity zone structure from a global analysis of SPdKS waves. *J. Geophys. Res.* **109,** B08301, doi:10.1029/2004JB003010 (2004).
26. Hernlund, J. W. & Jellinek, A. M. Dynamics and structure of a stirred partially molten ultralow velocity zone. *Earth Planet. Sci. Lett.* **296,** 1–8 (2010).
27. Hier-Majumder, S. Influence of contiguity on seismic velocities of partially molten aggregates. *J. Geophys. Res.* **113,** B12205, doi:10.1029/2008JB005662 (2008).
28. Fialin, M., Catillon, G. & Andrault, D. Disproportionation of Fe^{2+} in Al-free silicate perovskite in the laser heated diamond anvil cell as recorded by electron probe microanalysis of oxygen. *Phys. Chem. Miner.* **36,** 183–191 (2009).
29. Komabayashi, T. et al. High-temperature compression of ferropericlase and the effect of temperature on iron spin transition. *Earth Planet. Sci. Lett.* **297,** 691–699 (2010).
30. Ricolleau, A. et al. Density profile of pyrolite under the lower mantle conditions. *Geophys. Res. Lett.* **36,** L06302, doi:10.1029/2008GL036759 (2009).

Author Contributions R.N., H.O., S.T. and K.H. performed high-pressure experiments and chemical analyses. S.M. carried out the ELNES measurements. H.I. and N.H. supported the XES study at SPring-8. R.N., K.H. and J.H. wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.H. (kei@geo.titech.ac.jp).

METHODS

Melting experiments at high pressure. High-pressure and high-temperature conditions were generated using laser-heated diamond-anvil cell (DAC) techniques. $(Mg_{0.89}Fe_{0.11})_2SiO_4$ olivine from KLB-1 natural peridotite was used as a starting material[31]. The powdered sample was pressed into a disk with a typical thickness of 15–25 μm and was placed into a hole drilled at the centre of the rhenium gasket, which was pre-indented to a thickness of 25–40 μm. No additional metal absorber was used. Liquid argon was loaded cryogenically as a pressure medium. The sample was compressed with bevelled 120-μm and 150-μm, or flat 300-μm, culet diamond anvils, depending on a target pressure. Heating was performed with a fibre laser or a doughnut-mode Nd:YLF laser using a double-side heating technique that minimizes axial temperature gradients within the sample[32]. The laser spot size was about 20–30 μm on the sample.

The sample was heated for less than 1 s in order to avoid chemical segregation due to thermal diffusion under relatively large temperature gradient (see 'Attainment of chemical equilibrium' section in Supplementary Information). This heating duration was too short to allow accurate temperature measurement. Alternatively, we estimated the temperature in our experiment from the reported liquidus temperature of Mg_2SiO_4 (refs 8, 33) and the solidus temperature of natural peridotite[34], which give the upper and lower bounds, respectively (Supplementary Fig. 1). Pressure was determined from the Raman peak shift of the diamond anvil at 300 K after laser-heating[35]. Here we assumed that the pressure was higher by 20% at melting temperature than at 300 K due to a contribution of thermal pressure, which was estimated from the previous *in-situ* X-ray diffraction measurements[36]. The overall pressure uncertainty may be about ±10%.

Chemical analyses of recovered samples. After complete pressure release, the sample was recovered from the DAC and glued on a silicon wafer using polymeric resin. Subsequently it was polished parallel to the compression axis by an Ar ion beam using an Ion Slicer (JEOL EM-09100 IS)[37]. This method minimizes the damage of the sample surface, allowing us to observe microstructures within the extremely small sample. The section was examined by FE-EPMA (JEOL JXA-8500F) using an acceleration voltage of 10 kV and a beam current of 12 nA. We collected X-ray mappings for Si, Mg and Fe for each sample (Fig. 1 and Supplementary Figs 3 and 4). Chemical compositions of the quenched liquid pool and the neighbouring solid phase (ferropericlase, perovskite, or post-perovskite, depending on pressure) were obtained with defocused (5-μm size) and focused (<2-μm spatial resolution) beams, respectively (Supplementary Table 1). Synthetic wollastonite ($CaSiO_3$), periclase (MgO), and haematite (Fe_2O_3) were used as standards. In runs 21 and 24, however, the analyses of ferropericlase were contaminated by the melt pocket and were therefore estimated by subtracting the melt composition from the EPMA analyses until the SiO_2 content in ferropericlase becomes zero.

In addition, the samples in runs 17, 18 and 21 were further processed to a thin film by using an Ion Slicer and examined under a transmission electron microscope (TEM). For run 18, the $Fe^{3+}/(Fe^{2+} + Fe^{3+})$ ratio of the quenched melt pocket was determined by electron energy-loss near-edge structure (ELNES) spectroscopy measurements (Supplementary Fig. 5). The Fe $L_{2,3}$-edge ELNES spectra were collected by using a JEM-2100M with an Enfina1000 spectrometer at Nagoya University. The measurements were made with a dispersion of 0.1 eV per channel, a typical energy resolution of about 0.8 eV, and an integration time of 25–100 s. The spectra were recorded using a macroscript for the Gatan Digital Micrograph (which controls the spectrometer) that corrects for energy drifts during data accumulation. This avoids the problem of peak broadening that follows prolonged data accumulation[38]. The incident beam current was about 0.1 nA, and the fluence rate was about $50\,e\,Å^{-2}\,s^{-1}$. The quantitative determination of the $Fe^{3+}/(Fe^{2+} + Fe^{3+})$ ratio is based on the white line intensities at the Fe $L_{2,3}$-edge, following the method described in ref. 39. The validity of the present analytical procedures was checked on a couple of silicate glass standards, whose $Fe^{3+}/(Fe^{2+} + Fe^{3+})$ ratios had been previously determined by Mössbauer spectroscopy measurements to be 0.38 and 0.83 (ref. 40). Our ELNES analyses determined the $Fe^{3+}/(Fe^{2+} + Fe^{3+})$ ratios for these glass samples to be 0.40 ± 0.02 and 0.79 ± 0.09. A systematic change in $Fe^{3+}/(Fe^{2+} + Fe^{3+})$ ratios was not found with increasing electron beam irradiation time, up to 100 s.

X-ray emission spectroscopy measurements. High resolution X-ray emission spectra (XES) of the Fe Kβ line were obtained for $(Mg_{0.95}Fe_{0.05})SiO_3$ glass at the BL12XU beamline of SPring-8 (ref. 41). The glass sample was prepared by quenching melt from 1,931 K in a CO_2-H_2 gas-mixing furnace. The oxygen fugacity in the high-temperature furnace was controlled to be slightly above the iron–wüstite buffer, in which all Fe was present as ferrous iron. The FE-EPMA analyses confirmed that the sample was glass without any quench-crystals and had a chemical composition of $(Mg_{0.95}Fe_{0.05})SiO_3$ as a consequence of partial loss of iron into the Pt basket that held the sample in the furnace. X-ray diffraction measurements also showed the starting material to be amorphous.

XES measurements were made at room temperature in the DAC with increasing pressure from 8 to 85 GPa (Fig. 3). We used a cubic boron nitride (inner) + beryllium (outer) composite gasket. The glass powder was loaded into a hole in the gasket, which was pre-indented to about 80 μm in thickness. In order to excite the sample through diamond, monochromatic X-rays at 11 keV were chosen as the incident beam. The incident X-ray beam was focused on and collimated to 15-μm size at the sample. A 1-m Rowland circle type spectrometer, equipped with 1-m spherical bent Ge(620) analyser, was used to acquire the emission spectra. These spectra were collected through the gasket in order to avoid a loss of intensity due to absorption by the diamond. The spectra covered the energy range from 7,020 to 7,080 eV with a resolution of 0.5 eV.

31. Takahashi, E. Melting of a dry peridotite KLB-1 up to 14 GPa: implications on the origin of peridotitic upper mantle. *J. Geophys. Res.* **91,** 9367–9382 (1986).
32. Shen, G., Mao, H. K. & Hemley, R. J. in *Advanced Materials '96* 149–152 (Proc. 3rd NIRIM International Symposium on Advanced Materials, 1996).
33. Presnall, D. C., Weng, Y.-H., Milholland, C. S. & Walter, M. J. Liquidus phase relations in the system MgO–MgSiO₃ at pressures up to 25 GPa — constraints on crystallization of a molten Hadean mantle. *Phys. Earth Planet. Inter.* **107,** 83–95 (1998).
34. Fiquet, G. *et al.* Melting of peridotite to 140 gigapascals. *Science* **329,** 1516–1518 (2010).
35. Akahama, Y. & Kawamura, H. High pressure Raman spectroscopy of diamond anvils to 250 GPa: method for pressure determination in the multimegabar pressure range. *J. Appl. Phys.* **96,** 3748–3751 (2004).
36. Ozawa, H. *et al.* Experimental study of reaction between perovskite and molten iron to 146 GPa and implications for chemically distinct buoyant layer at the top of the core. *Phys. Chem. Miner.* **36,** 355–363 (2009).
37. Tateno, S., Sinmyo, R., Hirose, K. & Nishioka, H. The advanced ion-milling method for preparation of thin film using Ion Slicer: application to a sample recovered from diamond-anvil cell. *Rev. Sci. Instrum.* **80,** 013901 (2009).
38. Sasano, Y. & Muto, S. Energy-drift correction of electron energy-loss spectra from prolonged data accumulation of low SNR signals. *J. Electron Microsc.* **57,** 149–158 (2008).
39. van Aken, P. A., Liebscher, B. & Styrsa, V. J. Quantitative determination of iron oxidation states in minerals using Fe $L_{2,3}$-edge electron energy-loss near-edge structure spectroscopy. *Phys. Chem. Miner.* **25,** 323–327 (1998).
40. Jayasuriya, K., O'Neill, H. S. C., Berry, A. J. & Campbell, S. J. A Mössbauer study of the oxidation state of Fe in silicate melts. *Am. Mineral.* **89,** 1597–1609 (2004).
41. Jarrige, I., Cai, Y. Q., Ishii, H., Hiraoka, N. & Bleuzen, A. Thermally activated charge transfer in a Prussian blue derivative probed by resonant inelastic x-ray scattering. *Appl. Phys. Lett.* **93,** 054101 (2008).

LETTER

Structure of the spliceosomal U4 snRNP core domain and its implication for snRNP biogenesis

Adelaine K. W. Leung[1]†, Kiyoshi Nagai[1] & Jade Li[1]

The spliceosome is a dynamic macromolecular machine that assembles on pre-messenger RNA substrates and catalyses the excision of non-coding intervening sequences (introns)[1–3]. Four of the five major components of the spliceosome, U1, U2, U4 and U5 small nuclear ribonucleoproteins (snRNPs), contain seven Sm proteins (SmB/B′, SmD1, SmD2, SmD3, SmE, SmF and SmG) in common[4,5]. Following export of the U1, U2, U4 and U5 snRNAs to the cytoplasm[6,7], the seven Sm proteins, chaperoned by the survival of motor neurons (SMN) complex, assemble around a single-stranded, U-rich sequence called the Sm site in each small nuclear RNA (snRNA), to form the core domain of the respective snRNP particle[8,9]. Core domain formation is a prerequisite for re-import into the nucleus[10], where these snRNPs mature via addition of their particle-specific proteins. Here we present a crystal structure of the U4 snRNP core domain at 3.6 Å resolution, detailing how the Sm site heptad (AUUUUUG) binds inside the central hole of the heptameric ring of Sm proteins, interacting one-to-one with SmE–SmG–SmD3–SmB–SmD1–SmD2–SmF. An irregular backbone conformation of the Sm site sequence combined with the asymmetric structure of the heteromeric protein ring allows each base to interact in a distinct manner with four key residues at equivalent positions in the L3 and L5 loops of the Sm fold. A comparison of this structure with the U1 snRNP at 5.5 Å resolution[11,12] reveals snRNA-dependent structural changes outside the Sm fold, which may facilitate the binding of particle-specific proteins that are crucial to biogenesis of spliceosomal snRNPs.

Proteins in the Sm family are characterized by Sm1 and Sm2 motifs joined by a variable linker[13–15] (Supplementary Fig. 1). SmB/B′, SmD1 and SmD3 contain extended C termini, whereas SmD2 and SmE contain extended N termini. The Sm-fold consists of an N-terminal α-helix and a five-stranded antiparallel β-sheet containing the Sm motifs and folded upon itself[16] (Supplementary Fig. 2). The subunit interfaces in the SmD1–SmD2 and SmD3–SmB heterodimers suggest that the seven Sm proteins are assembled in a ring[16] in the snRNP core domain[17,18], and this has been confirmed by the crystal structure of the U1 snRNP at 5.5 Å resolution[11,12]. Crystal structure of *Archaeoglobus fulgidus* Lsm-1 homo-heptamer in complex with penta-uridylate showed how Lsm-1 provides U-specificity[19,20]. However, crystal structures of U1 snRNP at 5.5 or 4.4 Å resolution have left the side chain interactions between the Sm site and Sm proteins unresolved[11,12,21]. Hence, it remains unknown how the heteromeric Sm proteins combine to specifically recognize the Sm site sequence.

We solved the structure of the U4 snRNP core domain (Supplementary Table 1) assembled on a fragment of U4 snRNA (Supplementary Figs 3 and 4), which crystallized[22] with 12 copies in the asymmetric unit (Supplementary Fig. 5). The seven Sm proteins form a ring with a relatively flat face over which the N-terminal helices lie, and a tapered face carrying the L4 loops (Fig. 1a). U4 snRNA threads through the central hole lined by loops L3, L5 and L2 (Fig. 1c). The Sm site sequence is bound around the inner wall near the rim on the flat face. Its phosphates are exposed in the hole (Fig. 1b), revealing an irregular

backbone conformation (Fig. 2), and its bases project into the Sm proteins and are bound by their loops L3 and L5; they vary in orientation from nearly parallel to nearly perpendicular to the ring plane (Fig. 2 and Supplementary Fig. 6). The 5′ flanking adenine (A118) is outside the hole, and the 5′-stem (U4 SL-II) makes little contact with the Sm proteins (Fig. 1a). The 3′ flanking nucleotides (A126–C127) traverse the hole with phosphates contacting SmD1 and SmB, and bases contacting L2 of SmD2 and SmF and L5 of SmF. As the 3′-stem (U4 SL-III) emerges on the tapered face with the helical axis roughly 60° to the plane of the ring, the first base pair (C127:G144) comes into contact with Trp 25 of SmF (Trp 25F) in loop L2. The 3′ unpaired nucleotide (G145) is wedged between SmE and SmF. Along the 3′-stem the phosphates on both strands interact with basic residues from L2 and L4 loops of all the Sm proteins except SmG, particularly those from the lysine-rich, long L4 loops of SmD2 and SmB (Supplementary Fig. 7). These interactions can support the association between Sm proteins and different snRNAs during core domain assembly[23,24].

Small shape differences over the β-sheet of different Sm proteins (Supplementary Table 2), attributable to the size differences of the conserved inward pointing residues between the corresponding Sm1 and Sm2 motifs (Supplementary Fig. 1), cause the Sm protein ring to be asymmetric. The L3 and L5 loops are held at different heights and orientations relative to the plane of the ring (Supplementary Fig. 8), allowing their key residues to contact each nucleotide uniquely. The nucleotide-binding loops L3 and L5 have the consensus sequence of Asp-hydrophilic-aromatic-Met-Asn (residues L3.1–L3.5) and Arg-Gly-(acidic/Asn) (residues L5.1–L5.3) (Supplementary Table 3). In the co-crystal structure of *A. fulgidus* Lsm-1 with penta-uridylate, U specificity is achieved by sandwiching the uridine base between the side chains of His 37(L3.3) and Arg 63(L5.1), and hydrogen-bonding of its N3 and O4 atoms respectively with Oδ1 and Nδ2 of the invariant Asn 39(L3.5)[19]. In the U4 core domain (Fig. 3), however, the base stacking with the aromatic residue (L3.3) present in five of the Sm proteins, and the interaction with the Arg/Lys (L5.1), vary along the heptad and from the Lsm-1 example (Fig. 3 and Supplementary Fig. 9), due to the irregular RNA conformation around the central hole (Fig. 2 and Supplementary Fig. 6) and the sequence variations in L3 and L5 (Supplementary Table 3). We have inferred hydrogen-bonding interactions around the Sm site from the residue positions, corroborated by unambiguous orientation of base planes and aromatic side chains in the electron density, and the non-crystallographic symmetry (ncs) agreement in the conformation of the conserved contact residues (Fig. 3).

At the first position of the heptad an adenine is required, because replacement by G abolished core domain assembly with a U4 oligonucleotide, and replacement by U destabilized the assembly[23]. The A119 base is stacked with Tyr 53E(L3.3) and hydrogen-bonded at N3 to Lys 80E(L5.1), at N6 to Asp 51E(L3.1) and at N1 to Asn 55E(L3.5) (Fig. 3a), demonstrating A-specificity of its binding pocket. A salt bridge between its phosphate and Arg 61D2, which is hydrogen-bonded to
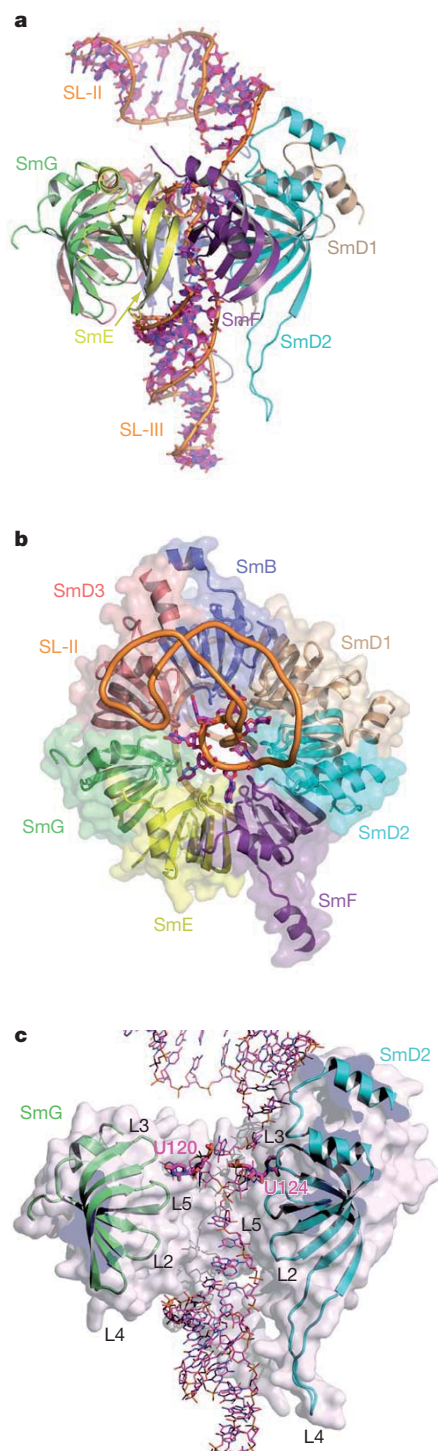
Figure 1 | Overall structure of the U4 snRNP core domain. a, Side view of the core domain showing the ring with its flat face up and tapered face down. **b**, View from the flat face of the ring. The N- and C-terminal extensions of the Sm fold interact between SmD3 and SmB, and between SmD1 and SmD2. **c**, The heptameric ring is cleaved along a plane (dark blue patches) through SmG and SmD2, leaving the five subunits SmG–SmD3–SmB–SmD1–SmD2 that bind the penta-uridylate to form the protein envelope in the background. Loops L3, L5 and L2 of the Sm fold line the walls of the funnel shaped hole, whereas L4 is exposed on the tapered face. The bases of the Sm site nucleotides, such as U120 and U124, are bound between L3 and L5 near the rim on the flat face.

Asp 37F (Fig. 3g), stabilizes A119 binding to the ring. In U1 snRNP and the reconstituted U4 core domain, the N7 atom of this adenine is unexpectedly nucleophilic in becoming methylated by dimethylsulphate. It indicates a perturbation of the π-electron distribution over
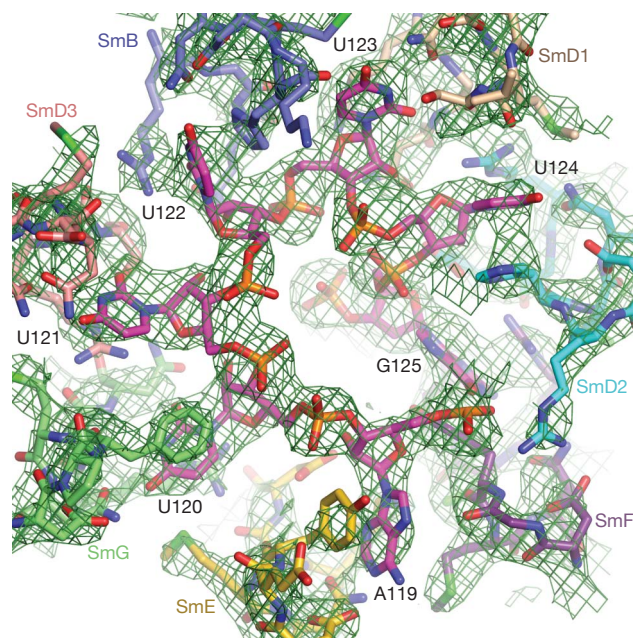

Figure 2 | The Sm site RNA binds asymmetrically in the central hole of the heptamer ring. The Sm site heptad sequence is shown together with the contact residues on Sm proteins around the central hole of the core domain. Carbon atoms are colour-coded by chain: SmE (yellow), SmG (green), SmD3 (salmon), SmB (light blue), SmD1 (tan), SmD2 (cyan), SmF (purple), RNA (magenta). Nitrogen, oxygen, sulphur and phosphorus atoms are in blue, red, green and orange, respectively. A sharpened ($B = -15$ Å^2) ncs-averaged electron density map is contoured at 8σ.

its double ring system[25], which could result from multiple hydrogen bonds to the base.

The U120 base is sandwiched between the side chains of Phe 37G and Arg 63G and hydrogen-bonded at O4 with Nδ2 of Asn 39G (Fig. 3b). Similarly, the U122 base is stacked with His 37B and hydrogen-bonded at O4 to Asn 39B (Fig. 3d), except that a different base orientation relative to L3 requires His 37B to adopt a rotamer different from that of His 62D2 (Supplementary Fig. 9). These interactions are similar to the Lsm-1 complex[19] and account for the cross-linking of the first and third U to L3 residues of SmG and SmB, respectively[26]. SmD3 and SmD1 lack the aromatic residue at L3.3. SmD3 displays U specificity in a novel mode distinct from the LSm-1 complex (ref. 19): U121 is hydrogen-bonded on O4 to both Asn 38D3 and Asn 40D3, on N3 to Asn 40D3 and on O2 to the peptide amide of Ser 66D3, besides being stacked with Arg 64D3 (Fig. 3c). U123 forms no stacking interaction (Fig. 3e). Its base is within hydrogen-bonding distance from the side chain of the invariant Asn 37D1, which is positioned to form the conserved buttressing hydrogen bonds with Asp 33D1[16,19,20]. This configuration is consistent with U123 adopting an enol tautomer that would present O2(H) and N3 as the hydrogen-bond donor and acceptor, respectively, to Oδ1 and Nδ2 of Asn 37D1. Consequently, U123 is accommodated without U-specific base contacts, which explains the lack of preference for U at this position (Supplementary Fig. 3e). In human U1 snRNA G replaces U, and in U1 snRNA of other species all four bases are tolerated (http://rfam.sanger.ac.uk)[27]. U124 is hydrogen-bonded on O2, not O4, to the invariant Asn 64D2 (Fig. 3f), and therefore our structure cannot fully account for the preference for U at this position. His 62D2 stacks with U124 and in some ncs copies also interacts edge-to-face with A118, the 5′-flanking adenine (not shown). The G125 base is not intimately associated with SmF, as it is stacked only on one edge between Tyr 39F and Arg 65F, and is too distant for hydrogen bonding with the invariant Asn 41F (Fig. 3g). Cys 66F, which is absolutely conserved in SmF, replaces the Gly in L5 without causing a clash. The absence of G-specific base contacts explains why replacing this G with A had no effect on Sm protein
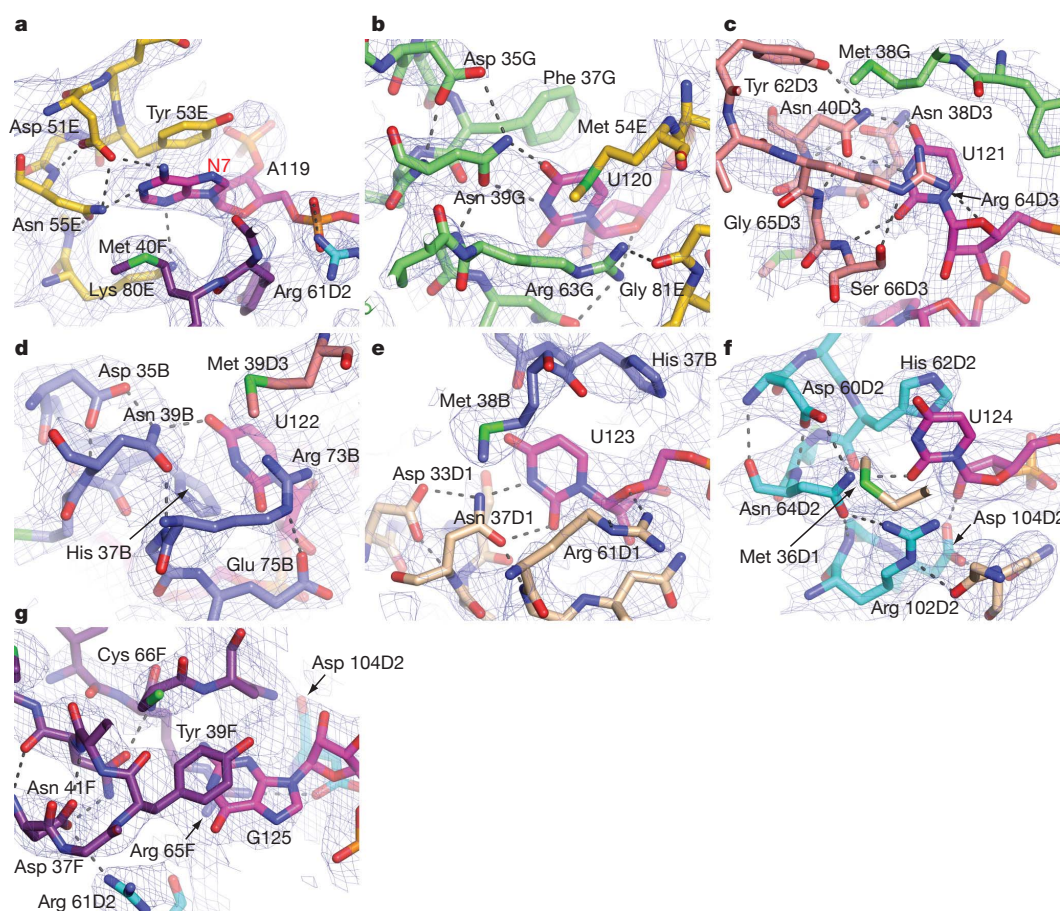
Figure 3 | Interactions between the U4 Sm site heptad nucleotides and the Sm proteins. a, A119. **b**, U120. **c**, U121. **d**,U122. **e**, U123. **f**, U124. **g**, G125. Shown in dashed lines are the hydrogen-bonding interactions inferred from the residue configurations. Similarities with the RNA-free heterodimers[16] and the Lsm-1 complex[19] are evident: the invariant Asn(L3.5) is buttressed by hydrogen bonds with the side chain of Asp(L3.1) and peptide amide of Gly/Cys(L5.2) in

six cases (**a**, **b**, **d–g**) and with Glu 36D3(L3.1) via Tyr 62D3 and with the peptide amide of Gly 65D3(L5.2) in SmD3 (**c**); Met(L3.4) contributes van der Waals contacts to the base bound by a neighbouring Sm protein in six cases (**a–f**), and Gly(L5.2) is conserved in six cases (**a–f**) where a side chain would clash with the base contacting residues. The $2F_o - F_c$ map is shown sharpened with $B = -15 \text{ Å}^2$ and contoured at ~1.5σ. Atom colours are as in Fig. 2.

binding[23], which is consistent with replacement of this G (Supplementary Fig. 3e) by other bases in U4 and U5 snRNAs of different species[27,28]. Thus, the last nucleotide of the Sm site heptad acts as a

transition to the variable 3′-stem. In the bound heptad the phosphate groups of U120, U121 and U122 are in close proximity. Their negative charge density is likely stabilized by electrostatically held magnesium
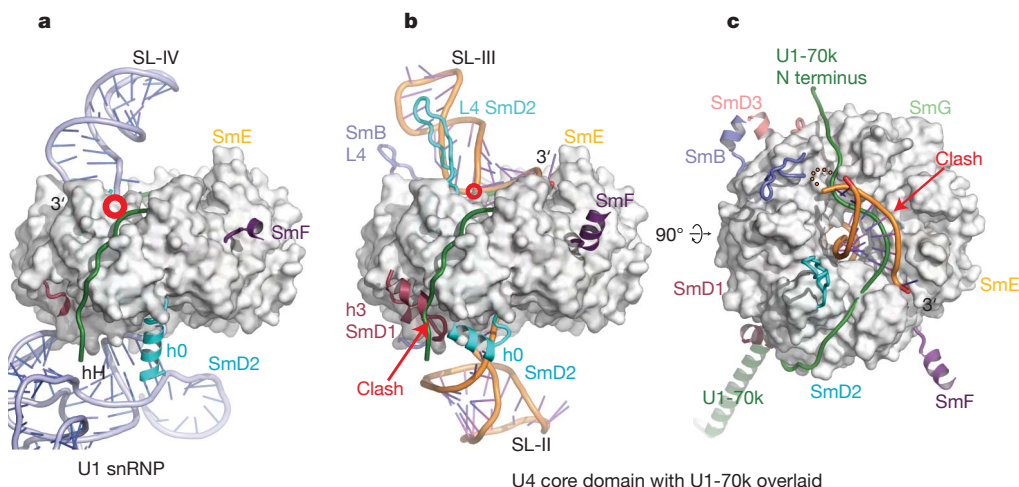
Figure 4 | snRNA-dependent structural changes of the U1 and U4 core domains. a, U1 snRNP[11] and **b**, U4 core domain in the same orientation; **c**, the U4 core domain in tapered-face view. The N-terminal fragment of U1-70K (green)[11,12] is overlaid onto the U4 core domain (**b** and **c**), and the Sm folds common to both structures are masked with a white envelope. The red circle indicates where U1 and U4 snRNAs emerge from the central hole. In U4, the L4

loops of SmB and SmD2 contact the backbone of the 3′-stem (**b**). In U1[11,21], SmD2 helix h0 points into the minor groove of RNA helix H (hH) (**a**), but in U4, SmD2 h0 is orientated almost orthogonal to this, with its N terminus pointing at SmD1 helix h3 (**b**). The latter is positioned to obstruct the path of U1-70K (arrow). Moreover the first seven base pairs of the 3′-stem of U4 snRNA (orange) would clash on its 3′-strand with U1-70K (arrow) (**c**).

ions of indeterminate positions[29] or by chelated cations unresolvable at our resolution.

In mammalian U4 and U5 snRNAs the Sm site heptad and the 3'-stem are linked by a single nucleotide, whereas in U1 and U2 snRNA they are linked by five nucleotides[27,28]. Structural comparison between U1 snRNP[11,12] and the U4 core domain shows that the snRNAs emerge from the central hole in similar positions (red circle in Fig. 4a and b) but with different stem orientations that cause their 3' termini to fall on opposite sides of the hole. In U1 snRNP, the N-terminal 60 residues of U1-70K wrap around the tapered face of the core domain by skirting around the RNA stalk passing SmD3–SmG–SmE–SmF–SmD2[11,12]. Mapping the U1-70K N-terminal fragment onto the U4 core domain shows that the 3' strand of U4 SL-III would obstruct its polypeptide path (Fig. 4c) and hence prevent its binding to the snRNP. The N-terminal 97-residue fragment of U1-70K is sufficient to bind to the U1 core domain, but fails to bind to the U5 core domain[30]. In U5 snRNA the 3'-stem is linked to the Sm site as closely as in U4 (Supplementary Fig. 3c, d) and could exclude U1-70K analogously.

The N-terminal extension of SmD2 and C-terminal extension of SmD1 are disordered in the absence of RNA[16]. In U1 snRNP, SmD2 forms an extra helix (helix 0) at the N terminus that points into the minor groove of RNA helix H (Fig. 4a). This anchors the SmD2 helix 1, whose C terminus interacts with U1-70K[11,12,21]. In the U4 core domain, the C-terminal extension of SmD1 forms helix 3, whereas the N-terminal extension of SmD2 forms helix 0 (Fig. 4b) in the ncs copies where the loop between SmD2 helices 0 and 1 interacts with the backbone of the 5'-stem (Fig. 1a). The SmD1 helix 3 interacting with the SmD2 helix 1 in its U4 snRNA-dependent orientation might also obstruct U1-70K (Fig. 4b). In U1 snRNP, the SmB helix 1 interacts with the backbone of stem II (ref. 11). These snRNA-dependent structural differences on the flat face of the core domain may, in addition to the snRNA itself, provide selectivity for the cognate particle-specific proteins and have a crucial role in snRNP biogenesis.

METHODS SUMMARY

The U4 snRNP core domain was reconstituted from the seven Sm proteins and a variant of human U4 snRNA[16,22]. Crystals in space group $P3_1$ with 12 complexes per asymmetric unit were grown by vapour diffusion and diffracted X-rays anisotropically to 3.45 Å resolution. Initial phases were determined by the multiwavelength anomalous diffraction (MAD) method using SeMet substitution within Sm sub-complexes. The structure containing 8,101 protein and RNA residues has been refined under 12-fold ncs restraints at 66.2–3.6 Å resolution to R_{free} of 32.1% with excellent geometry (Supplementary Table 1).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Burge, C. B., Tuschl, T. & Sharp, P. A. in *The RNA World* 2nd edn (eds Gesteland, R. R., Cech, T. R. & Atkins, J. F.) 525–560 (Cold Spring Harbor Laboratory Press, 1999).
2. Will, C. L. & Lührmann, R. in *The RNA World* 3rd edn (eds Gesteland, R. F., Cech, T. R. & Atkins, J. F.) 369–400 (Cold Spring Harbor Laboratory Press, 2006).
3. Yu, Y.-T., Scharl, E. C., Smith, C. M. & Steitz, J. A. in *The RNA World* 2nd edn (eds Gesteland, R. R., Cech, T. R. & Atkins, J. F.) 487–524 (Cold Spring Harbor Laboratory Press, 1999).
4. Hinterberger, M., Pettersson, I. & Steitz, J. A. Isolation of small nuclear ribonucleoproteins containing U1, U2, U4, U5, and U6 RNAs. *J. Biol. Chem.* **258**, 2604–2613 (1983).
5. Bringman, P. & Lührmann, R. Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents. *EMBO J.* **5**, 3509–3516 (1986).
6. Mattaj, I. W. Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell* **46**, 905–911 (1986).
7. Ohno, M., Segref, A., Bachi, A., Wilm, M. & Mattaj, I. W. PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. *Cell* **101**, 187–198 (2000).
8. Meister, G., Eggert, C. & Fischer, U. SMN-mediated assembly of RNPs: a complex story. *Trends Cell Biol.* **12**, 472–478 (2002).
9. Pellizzoni, L., Yong, J. & Dreyfuss, G. Essential role for the SMN complex in the specificity of snRNP assembly. *Science* **298**, 1775–1779 (2002).
10. Fischer, U., Sumpter, V., Sekine, M., Satoh, T. & Lührmann, R. Nucleo-cytoplasmic transport of U snRNPs: definition of a nuclear location signal in the Sm core domain that binds a transport receptor independently of the m3G cap. *EMBO J.* **12**, 573–583 (1993).
11. Pomeranz Krummel, D. A., Oubridge, C., Leung, A. K. W., Li, J. & Nagai, K. Crystal structure of the human spliceosomal U1 snRNP at 5.5 Å resolution. *Nature* **458**, 475–480 (2009).
12. Oubridge, C., Pomeranz Krummel, D. A., Leung, A. K. W., Li, J. & Nagai, K. Interpreting a low resolution map of human U1 snRNP using anomalous scatterers. *Structure* **17**, 930–938 (2009).
13. Hermann, H. *et al.* snRNP Sm proteins share two evolutionarily conserved sequence motifs which are involved in Sm protein–protein interactions. *EMBO J.* **14**, 2076–2088 (1995).
14. Seraphin, B. Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *EMBO J.* **14**, 2089–2098 (1995).
15. Cooper, M., Johnston, L. H. & Beggs, J. D. Identification and characterization of Uss1p (Sdb23p): a novel U6 snRNA-associated protein with significant similarity to core proteins of small nuclear ribonucleoproteins. *EMBO J.* **14**, 2066–2075 (1995).
16. Kambach, C. *et al.* Crystal structures of two Sm protein complexes and their implications for the assembly of the spliceosomal snRNPs. *Cell* **96**, 375–387 (1999).
17. Kastner, B. & Lührmann, R. Electron microscopy of U1 small nuclear ribonucleoprotein particles: shape of the particle and position of the 5' RNA terminus. *EMBO J.* **8**, 277–286 (1989).
18. Kastner, B., Bach, M. & Lührmann, R. Electron microscopy of small nuclear ribonucleoprotein (snRNP) particles U2 and U5: evidence for a common structure-determining principle in the major U snRNP family. *Proc. Natl Acad. Sci. USA* **87**, 1710–1714 (1990).
19. Törö, I. *et al.* RNA binding in an Sm core domain: X-ray structure and functional analysis of an archaeal Sm protein complex. *EMBO J.* **20**, 2293–2303 (2001).
20. Thore, S., Mayer, C., Sauter, C., Weeks, S. & Suck, D. Crystal structures of the *Pyrococcus abyssi* Sm core and its complex with RNA. Common features of RNA binding in Archaea and Eukarya. *J. Biol. Chem.* **278**, 1239–1247 (2003).
21. Weber, G., Trowitzsch, S., Kastner, B., Lührmann, R. & Wahl, M. C. Functional organization of the Sm core in the crystal structure of human U1 snRNP. *EMBO J.* **29**, 4172–4184 (2010).
22. Leung, A. K. W. *et al.* Use of RNA tertiary interaction modules for the crystallization of the spliceosomal snRNP core domain. *J. Mol. Biol.* **402**, 154–164 (2010).
23. Raker, V. A., Hartmuth, K., Kastner, B. & Lührmann, R. Spliceosomal U snRNP core assembly: Sm proteins assemble onto an Sm site RNA nonanucleotide in a specific and thermodynamically stable manner. *Mol. Cell. Biol.* **19**, 6554–6565 (1999).
24. McConnell, T. S., Lokken, R. P. & Steitz, J. A. Assembly of the U1 snRNP involves interactions with the backbone of the terminal stem of U1 snRNA. *RNA* **9**, 193–201 (2003).
25. Hartmuth, K., Raker, V. A., Huber, J., Branlant, C. & Lührmann, R. An unusual chemical reactivity of Sm site adenosines strongly correlates with proper assembly of core U snRNP particles. *J. Mol. Biol.* **285**, 133–147 (1999).
26. Urlaub, H., Raker, V. A., Kostka, S. & Lührmann, R. Sm protein–Sm site RNA interactions within the inner ring of the spliceosomal snRNP core structure. *EMBO J.* **20**, 187–196 (2001).
27. Gardner, P. P. *et al.* Rfam: updates to the RNA families database. *Nucleic Acids Res.* **37**, D137–D140 (2009).
28. Guthrie, C. & Patterson, B. Spliceosomal snRNAs. *RNA* **9**, 193–201 (2003).
29. Draper, D. E. RNA folding: thermodynamic and molecular descriptions of the roles of ions. *Biophys. J.* **95**, 5489–5495 (2008).
30. Nelissen, R. L., Will, C. L., van Venrooij, W. J. & Lührmann, R. The association of the U1-specific 70K and C proteins with U1 snRNPs is mediated in part by common U snRNP proteins. *EMBO J.* **13**, 4113–4125 (1994).

Author Contributions A.K.W.L. and K.N. designed the constructs. A.K.W.L. crystallized the core domain, collected data and solved the structure in $P6_122$. J.L. identified twinning and refined the structure in $P3_1$. All three authors wrote the paper.

Author Information Atomic coordinates and structure factors for the U4 snRNP core domain have been deposited in the PDB data bank under accession numbers 2Y9A, 2Y9B, 2Y9C and 2Y9D. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.N. (kn@mrc-lmb.cam.ac.uk) and J.L. (jl@mrc-lmb.cam.ac.uk).

METHODS

Crystallization. Human Sm proteins were overproduced as two heterodimers, SmD1–SmD2 and SmD3–SmB, and a trimer, SmE–SmF–SmG, in *Escherichia coli* as described previously[16,22]. The Sm proteins expressed were full-length except for SmB, whose extended C terminus was truncated to residue 95. The 3′ fragment of human U4 snRNA from residues 85–145, comprising the Sm site and both flanking helices (Supplementary Fig. 4a), was produced by *in vitro* transcription[22,31]. As the U4 Sm site is almost immediately flanked by helices (Supplementary Fig. 4a), this fragment was thought likely to form a compact assembly with the Sm proteins. To promote crystal contacts, we replaced residues 97–104 of the native sequence by a stable GAAA tetraloop and residues 134–137 by a GAAA tetraloop receptor (Supplementary Fig. 4b). The human U4 core domain was reconstituted from the Sm proteins and the modified 3′ fragment of U4 snRNA, and purified on a monoQ column. Crystals were obtained by sitting drop vapour diffusion using a reservoir solution containing 4–10% (v/v) PEG 550 MME, 0.1 M KSCN or $(NH_4)_2SO_4$, 0.1 M Tris-HCl pH 8.0–8.5, 10 mM MgCl$_2$, and 2 mM cyclen (a polyamine)[32]. A full account of our crystallization effort has been published elsewhere[22].

Preparation of Se-Met derivatives. Se-Met labelled SmD1–SmD2 and SmD3–SmB heterodimers were expressed by the inhibition method[33]. Se-Met labelled SmE–SmF–SmG hetero-trimer was expressed in the methionine auxotroph B834, grown with the Se-Met core medium (Wako, 391-01541) as supplement. Incorporation of Se-Met was verified by electrospray time-of-flight mass spectrometry of the purified proteins. By replacing one or all of the heterodimers and trimer with the Se-Met-labelled subcomplexes during the reconstitution, we obtained differently labelled Se-Met derivatives of the core domain.

Data processing. Data from Se-Met derivatives were processed using MOSFLM[34], with anisotropic resolution limits when required, SCALA[35] and TRUNCATE[36]. Native data were integrated using XDS[37], converted to the CCP4 format using POINTLESS[35], and scaled using SCALA. Between randomly partitioned half-sets of native data the correlation[35] was 92% at 4.1 Å, decreasing to ~50% at 3.6 Å, and therefore data to 3.6 Å were included in the refinement (Supplementary Table 1). Before refinement, the native amplitudes were corrected for anisotropy with truncation in the weak direction to $F/\sigma F \geq 3$ (www.doe-mbi.ucla.edu/~sawaya/anisoscale)[38].

The crystals are of space group $P3_1$ with 12 copies of the core domain in the asymmetric unit related by 222 rotational non-crystallographic symmetry (ncs) and threefold translational non-crystallographic symmetry in the *a*,*b*-plane (translational pseudo-symmetry). The crystals are twinned along the *a*,*b*- and *a**,*b**-axes, but twinning was masked by the pseudo-symmetry and anisotropy, resulting in apparent $P6_122$ symmetry for data sets at low resolution. The pseudo-symmetry gives rise to self-Patterson vectors at (1/3, 2/3, 0) and (2/3, 1/3, 0), and causes intensity modulations up to 5 Å resolution where two-thirds of the rows, having indices of $(h − k \neq 3n, h + 2k \neq 3n)$, are systematically weak. The other rows of strong reflections can be selected as those retaining an integral index after re-indexing *hkl* to $(h/3 − k/3, h/3 + 2k/3, l)$, and they correspond to a small unit cell whose a, b-edges are reduced by a factor of $\sqrt{3}$.

Initial phasing in the small cell. Initial phases were obtained at 5.5 Å resolution by the MAD method[39] using a crystal labelled with Se-Met in the SmE–SmF–SmG proteins (SeEFG-MAD, Supplementary Table 1). The phases were calculated in the apparent space group of $P6_122$ and for the rows of strong reflections only. This procedure approximates both the rotational and translational ncs as crystallographic, and yields phases for the small unit cell ($a = 142.1$ Å, $c = 146.1$ Å for SeEFG-MAD) in which the asymmetric unit contains only one core domain representing the average of ncs copies with structural and orientational differences.

Fourteen Se sites were found using SHELXD[40] from the anomalous signal of the peak wavelength data of SeEFG-MAD. The Se positions were refined and phases calculated in SHARP[41]. The MAD phases showed an overall figure-of-merit of 0.649 (acentric) and 0.495 (centric) at 5.5 Å resolution for 12 sites. The anomalous phasing power was 1.742, 1.033 and 1.030 for the peak, inflection and remote wavelengths; the isomorphous phasing power was 0.913 (acentric) and 0.713 (centric) for the inflection, and 1.060 (acentric) and 0.860 (centric) for the remote. Using solvent-flattened phases from RESOLVE[42,43] and peak wavelength data to 5 Å resolution, the hetero-heptameric ring model for the core domain[16] was located in the density by a phased molecular replacement using MOLREP[44] in the automatic rotation-translation search mode. The search model consisted of a superposition of the ring model from ref. 16 and the homo-heptameric ring from the Lsm crystal structure (PDB ID 1M8V)[20]. The top solution showed Met residues in SmE–SmF–SmG proteins of the model overlapping with the majority of the Se anomalous peaks. The SeEFG-MAD phases were used to locate Se sites in the small cell, in anomalous difference maps calculated from the Se peak wavelength data from crystals containing Se-Met labels in other combinations of Sm proteins.

Partial model in the small cell. The SeEFG-MAD map of the small cell, even before phase improvement, showed clear density for an RNA helix later identified as part of the 3′-stem (Supplementary Fig. 10a). Presence for the bound Sm site RNA could be inferred from the oblong shape of the central hole of the ring. Density for the 5′-stem was poor but weakly connected to the Sm site region. There is density for the long L4 loop of SmD2, which was disordered in the SmD1–SmD2 heterodimer structure[16], but not for L4 of SmB, which was ordered in the SmD3–SmB heterodimer[16]. The regions showing poor densities at this stage were later found to exhibit greater variations among the ncs copies. A partial model containing the seven Sm proteins and a fragment of the 3′ RNA helix was built into the density in the small cell of the SeEFG-MAD derivative, using the program O[45] (Supplementary Fig. 10b).

Experimental map in the true cell. The partial model built in the small cell of SeEFG-MAD was used as the search model for molecular replacement by AMoRe[46] to the large cell (true cell) of each derivative in $P6_122$. Resolution for the rotation search was 7 Å; resolution for the translation searches was 4–6 Å for the first and second copies but reduced to 5–7.5 Å for the third copy as necessary. Acceptance of the translation solution was based on a minimal separation of 62 Å on finding the first copy, and displacement vectors between copies that agree with the self-Patterson. For all derivatives the solution after rigid-body refinement showed correlation with F_{obs} of >54% at 5.5 Å resolution. Thus the pseudo-translation was resolved.

The partial model for SeEFG-MAD was also mapped into the small cell of the other, non-isomorphous derivatives by molecular replacement in MOLREP[44]. For each derivative, the matrices relating the partial model in the small cell to its copies in the large cell were determined by the lsq_explicit command in O[45] and used to copy the Se sites into the large cell. Phases from different derivatives in the large cell were combined by multi-domain, multi-crystal averaging in DMMULTI[47] at 4.5 Å resolution, using three separate averaging masks over the ring, 3′- and 5′-stems of the partial model. Using the DMMULTI[47] phases with SeEFG-SAD amplitudes, we calculated an experimental map in the large $P6_122$ cell that allowed the building of the complete RNA.

Initial model in the large cell. The heptameric rings were rebuilt in the large cell. The Cα-trace of a homo-heptameric Lsm ring (PDB ID 1M8V)[20] was superimposed on the partial model by the Sm1, Sm2 motifs, to provide a regular scaffold for the seven inter-subunit β-sheets[16] that is a conserved characteristic of the Sm family. The crystal structures of SmD1, SmD2, SmD3 and SmB[16] were aligned to subunits in the 1M8V scaffold and adjusted to the density. For SmE, SmF and SmG, a composite template, consisting of the four Sm protein structures (PDB ID 1D3B, 1B34)[16], two Lsm structures (PDB ID 1LJO, 1M8V)[20,48], and the yeast SmF structure (PDB ID 1N9R)[49] in superposition, was aligned to the scaffold, and the model was rebuilt by modifying parts of the composite template that fitted the density better than the small-cell model. Additional adjustment was necessary to bring the Met side chains near the Se anomalous peaks. In SmF significant shifts of the β1 and β2 strands from the small-cell model were required to fit Se peaks for Met 20F and Met 27F.

Rigid-body refinement of sheet tilt. The model was subjected to refinement alternated with rebuilding using O[45] and Coot[50]. Starting from an *R*-factor of 56% at 4 Å resolution, rigid-body refinement was done in CNS[51] against the mlhl target of the DMMULTI[47] phases and SeEFG-SAD amplitudes, using one rigid group per core domain and then three groups of the ring, 5′- and 3′-stems. The resulting $2F_o − F_c$ map allowed the flanking RNA stems to be built in different conformations for the three ncs copies. Making each Sm protein a separate rigid group did not lower R_{free}, but making each inter-subunit β-sheet into a rigid group, involving cutting each protein across the middle of strands β2–β4, did reduce the R_{free} further by 2%. Therefore the seven inter-subunit β-sheets are found to have different tilts relative to the pseudo-sevenfold axis of the ring. Although the differences are moderate, the resulting architecture of the ring places the RNA binding loops (L3 and L5) in each Sm protein at different heights (Supplementary Fig. 8e, f) relative to the ring axis, thus predisposing the core ring to recognize RNA in circularly asymmetrical conformation.

Minimisation in $P6_122$. The model was further refined by minimisation in CNS at 3.6 Å resolution against the mlhl target of SeEFG-SAD amplitudes with DMMULTI[47] phases. The ncs restraints were applied in a single equivalence set that contained the Sm1, Sm2 motifs of all seven proteins and the Sm site RNA from the same ring. Additional restraints were imposed on base pairing in the RNA helices and protein main-chain hydrogen bond distances (2.9 ± 0.25 Å, 50 kcal mol^{-1} Å^{-2}). The R_{free} decreased to ~50%. Re-determining the Se sites using the model phase reduced the R_{free} of minimisation to ~48%, whereupon the improved $2F_o − F_c$ map allowed rebuilding of the Sm site region and the addition of the L4 loops of SmB as well as the N-, C-terminal extensions from the Sm-fold. When a fully Se-Met substituted derivative (Se-all, Supplementary Table 1) became available, the model built for the SeEFG-SAD crystal was positioned in that cell to calculate an

anomalous difference map at 5 Å resolution, which showed Se peaks accompanying each unique Se-Met residue in at least one copy. These peaks helped to place secondary structures in the extensions from the Sm fold. The SeEFG-SAD amplitudes were then replaced by the native, which reduced the R_{free} to ~46%.

Twinning. After re-indexing the native data in subgroups of $P6_122$, the self-rotation peaks at $\omega = 90°$, $\kappa = 180°$ fell with increasing resolution, indicating that the twofold rotation symmetries perpendicular to the c-axis are non-crystallographic. Analysis of the native intensities showed twinning about the a,b- and a^*,b^*-axes. The space group was re-classified as $P3_1$, and the model was duplicated by these twofold operators to generate six, then twelve copies per asymmetric unit, which were subjected to rigid-body refinement, giving R_{free} of 44.5%. B-factors were refined using data at 10–3.6 Å resolution. However, twin refinement by minimisation and bgroup using CNS, which assumes merohedral twinning, could not reduce R_{free} below 44%.

Twin refinement in $P3_1$. Refmac5 (refs 52, 53) found the crystal to be perfectly twinned with four twin domains of ~25% twin fractions. Correcting the representation of twinning immediately lowered R_{free} by ~10%. Rigid-body refinement with one core domain per rigid group against the anisotropy-corrected amplitudes at 20–5 Å reduced the R_{free} by 1.2% over 30 cycles. Further rigid-body refinement with one chain of protein or RNA per group caused some chains to rotate by more than 3° and reduced R_{free} by another 0.4%, to 32.6% at 5 Å resolution.

Three rounds of individual atom refinement in Refmac5 followed, with extensive rebuilding in-between. In the first round, 12-fold ncs restraints were still applied in a single equivalence set containing the Sm1, Sm2 motifs of all seven proteins and the Sm site RNA within one complex. In the later rounds, eight equivalence sets were defined, seven for the Sm1, Sm2 motifs of the seven proteins and one for the Sm site RNA. Main-chain hydrogen-bond restraints of 2.9 ± 0.2 Å were applied to the proteins throughout, using a list output by hydrogen_bonds.inp in CNS[51] and converted to the Refmac5 (ref. 52) format. No external restraints were imposed on the RNA. Ncs-averaged maps were calculated in Coot[50] from the $2F_o - F_c$ map, where the ncs matrices were evaluated between whole complexes using coordinates truncated to contain only the Sm1, Sm2 motifs and the Sm site RNA. The averaged map was used for rebuilding in the initial round wherever the $2F_o - F_c$ map was unclear; but in later rounds the $2F_o - F_c$ map was increasingly relied upon.

The first round of individual atom refinement was run for only one cycle to generate the $2F_o - F_c$ map at 3.6 Å resolution, which revealed that loop L1 of SmE, previously built to be like L1 of SmD2, was two residues too long and should be like L1 of the remaining Sm proteins. The other significant change was in the β-sheet of SmD1. Loop L1 forms the SmD1–SmD1 interface across an ncs twofold. After the rigid body refinement, at four of these interfaces there is a potential clash, which was relieved by shifting the β-strands towards β4 (see Supplementary Fig. 2) in accordance with the $2F_o - F_c$ map. This accounts for the greater root-mean-square Cα distances[54] among ncs copies of SmD1 compared with the other Sm proteins (Supplementary Table 2). In view of the modest resolution, Ramachandran restraints were turned on during real-space refinement in Coot.

The final model contains 7,285 protein, and 816 RNA residues. It has been refined at 66.2–3.6 Å resolution against 158,528 reflections (82.9% completeness) to $R = 27.6\%$, $R_{free} = 32.1\%$. An $F_o - F_c$ omit map of the Sm site, calculated by repeating the final refinement with A119–G125 omitted in all ncs copies, confirmed the conformation of the heptad (Supplementary Fig. 11). According to MolProbity[55], the model shows good stereochemistry in the 89th percentile of structures in the resolution range of 3.25–3.85 Å, with 91.8% of the protein residues in the favoured regions of the Ramachandran plot. All molecular structure figures were drawn using PyMOL[56].

31. Price, S. R., Ito, N., Oubridge, C., Avis, J. M. & Nagai, K. Crystallization of RNA–protein complexes. I. Methods for the large-scale preparation of RNA suitable for crystallographic studies. *J. Mol. Biol.* **249,** 398–408 (1995).
32. Sauter, C. *et al.* Additives for the crystallization of proteins and nucleic acids. *J. Cryst. Growth* **196,** 365–376 (1999).
33. Doublié, S. Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276,** 523–530 (1997).
34. Leslie, A. G. W. The integration of macromolecular diffraction data. *Acta Crystallogr. D* **62,** 48–57 (2006).
35. Evans, P. R. Scaling and assessment of data quality. *Acta Crystallogr. D* **62,** 72–82 (2006).
36. French, G. S. & Wilson, K. S. On the treatment of negative intensity observations. *Acta Crystallogr. A* **34,** 517–525 (1978).
37. Kabsch, W. XDS. *Acta Crystallogr. D* **66,** 125–132 (2010) CrossRef.
38. Strong, M. *et al.* Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis. Proc. Natl Acad. Sci. USA* **103,** 8060–8065 (2006).
39. Hendrickson, W. A. & Ogata, C. M. Phase determination from multiwavelength anomalous diffraction measurements. *Methods Enzymol.* **276,** 494–523 (1997).
40. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* **58,** 1772–1779 (2002).
41. de La Fortelle, E. & Bricogne, G. Maximum likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276,** 472–494 (1997).
42. Terwilliger, T. C. SOLVE and RESOLVE: automated structure solution and density modification. *Methods Enzymol.* **374,** 22–37 (2003).
43. Terwilliger, T. C. SOLVE and RESOLVE: automated structure solution, density modification and model building. *J. Synchrotron Radiat.* **11,** 49–52 (2004).
44. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30,** 1022–1025 (1997).
45. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47,** 110–119 (1991).
46. Navaza, J. Implementation of molecular replacement in AmoRe. *Acta Crystallogr. D* **57,** 1367–1372 (2001).
47. Cowtan, K. D., Zhang, K. Y. J. & Main, P. In *International Tables for Crystallography, Volume F. Crystallography of Biological Macromolecules* (eds. Rossmann, M. G. & Arnold, E.) 705–710 (Kluwer Academic Publishers, 2001).
48. Törö, I., Basquin, J., Teo-Dreher, H. & Suck, D. Archaeal Sm proteins form heptameric and hexameric complexes: crystal structures of the Sm1 and Sm2 proteins from the hyperthermophile *Archaeoglobus fulgidus. J. Mol. Biol.* **320,** 129–142 (2002).
49. Collins, B. M. *et al.* Homomeric ring assemblies of eukaryotic Sm proteins have affinity for both RNA and DNA. Crystal structure of an oligomeric complex of yeast SmF. *J. Biol. Chem.* **278,** 17291–17298 (2003).
50. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66,** 486–501 (2010).
51. Brunger, A. T. Version 1.2 of the Crystallography and NMR system. *Nature Protocols* **2,** 2728–2733 (2007).
52. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53,** 240–255 (1997).
53. Vagin, A. *et al.* Organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D* **60,** 2184–2195 (2004).
54. Diamond, R. On the multiple simultaneous superposition of molecular structures by rigid body transformations. *Protein Sci.* **1,** 1279–1287 (1992).
55. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35,** W375–W383 (2007).
56. DeLano, W. L. The PyMOL Molecular Graphics System ⟨http://www.pymol.org⟩ (2002).

LETTER

Substrate–modulated gating dynamics in a Na$^+$-coupled neurotransmitter transporter homologue

Yongfang Zhao[1,2,4]*, Daniel S. Terry[5]*, Lei Shi[5,6]*, Matthias Quick[1,2,4], Harel Weinstein[5,6], Scott C. Blanchard[5] & Jonathan A. Javitch[1,2,3,4]

Neurotransmitter/Na$^+$ symporters (NSSs) terminate neuronal signalling by recapturing neurotransmitter released into the synapse in a co-transport (symport) mechanism driven by the Na$^+$ electrochemical gradient[1–6]. NSSs for dopamine, noradrenaline and serotonin are targeted by the psychostimulants cocaine and amphetamine[1], as well as by antidepressants[7]. The crystal structure of LeuT, a prokaryotic NSS homologue, revealed an occluded conformation in which a leucine (Leu) and two Na$^+$ are bound deep within the protein[8]. This structure has been the basis for extensive structural and computational exploration of the functional mechanisms of proteins with a LeuT-like fold[9–22]. Subsequently, an 'outward-open' conformation was determined in the presence of the inhibitor tryptophan[23], and the Na$^+$-dependent formation of a dynamic outward-facing intermediate was identified using electron paramagnetic resonance spectroscopy[24]. In addition, single-molecule fluorescence resonance energy transfer imaging has been used to reveal reversible transitions to an inward-open LeuT conformation, which involve the movement of transmembrane helix TM1a away from the transmembrane helical bundle[22]. We investigated how substrate binding is coupled to structural transitions in LeuT during Na$^+$-coupled transport. Here we report a process whereby substrate binding from the extracellular side of LeuT facilitates intracellular gate opening and substrate release at the intracellular face of the protein. In the presence of alanine, a substrate that is transported $\sim$10-fold faster than leucine[15,25], we observed alanine-induced dynamics in the intracellular gate region of LeuT that directly correlate with transport efficiency. Collectively, our data reveal functionally relevant and previously hidden aspects of the NSS transport mechanism that emphasize the functional importance of a second substrate (S2) binding site within the extracellular vestibule[15,20]. Substrate binding in this S2 site appears to act cooperatively with the primary substrate (S1) binding site to control intracellular gating more than 30 Å away, in a manner that allows the Na$^+$ gradient to power the transport mechanism.

The experiments were performed on LeuT engineered to contain a 15-amino-acid, carboxy-terminal biotinylation domain[26] and site-specifically labelled with the fluorophores Cy3 and Cy5 maleimide at residue position 7, after replacing the native His residue with Cys (H7C) in the amino-terminal loop close to TM1, and at position 86 (R86C) in intracellular loop (IL) 1 (Methods). Direct observations of conformational processes within the intracellular gate region of LeuT (Supplementary Fig. 1) were made using a wide field imaging strategy employing prism-based total internal reflection (Methods, Fig. 1a). As described[22], fluorescence resonance energy transfer (FRET) imaging of

LeuT revealed two readily distinguished states (FRET efficiency $\sim$0.51 and $\sim$0.75) in the presence of 200 mM K$^+$ and the nominal absence of Na$^+$ (Fig. 1b), consistent with the existence of two distinct conformations of the intracellular gate that differ by $\sim$13 Å in the distance separating the fluorophore pair.

In experiments imaging LeuT dynamics with increasing Na$^+$ concentrations, Hidden Markov Modelling revealed that the distribution of low- and high-FRET conformations of LeuT was altered by Na$^+$ with an effector concentration for half-maximum response (EC$_{50}$) of 10.9 mM (Fig. 1b, c), consistent with the EC$_{50}$ for Na$^+$-dependent stimulation of substrate binding and transport[15]. Na$^+$ decreased the overall frequency of transitions (Fig. 1d, e) through the preferential stabilization ($\sim$7-fold) of the inward-closed state. During the direct imaging of individual LeuT molecules (Fig. 1f), slow, spontaneous transitions between open and closed states, initially observed in 200 mM K$^+$, were dramatically decreased on exchange into Na$^+$-containing buffer, leading to the preferential stabilization of the inward-closed state.

Reasoning that substrate-induced intracellular gating might be observed best under conditions mimicking the relatively low intracellular Na$^+$, we performed experiments at Na$^+$ concentrations sufficient for Leu binding but below the EC$_{50}$ of Na$^+$. However, even at 2 mM Na$^+$, Leu shifted the population towards the closed intracellular gate conformation (Supplementary Fig. 2a, b) through a $\sim$3.5-fold stabilization of this state (Supplementary Fig. 2c). These effects, which result in a global decrease in transition frequency (Supplementary Fig. 2d), were recapitulated at the level of individual LeuT molecules (Supplementary Fig. 2e). Thus, while unambiguously demonstrating binding of both Na$^+$ and Leu to LeuT, these results corroborate our earlier finding that Leu binding has the net effect of diminishing the likelihood of intracellular gate opening. One possible explanation for these observations is that Leu's high affinity for the transporter[15] makes it a poor substrate for transport, which in our measurements is manifested in the greatly extended lifetime of the closed state. To test this hypothesis, intracellular gate dynamics were assessed in the presence of the more efficiently transported substrate Ala.

In stark contrast to Leu, under otherwise identical conditions, increasing Ala concentrations did not shift the FRET distribution towards the closed state (Fig. 2a, b). Instead, a strong, Ala-concentration-dependent enhancement of transition rates was observed. In 2 mM Na$^+$, Ala enhanced the transition rates between inward-open and inward-closed states by as much as $\sim$4-fold (Fig. 2c, d). This result was directly confirmed at the scale of individual molecules on exchange into Ala-containing buffer (Fig. 2e). Similar enhancements in transition frequency were also

[1]Center for Molecular Recognition, Columbia University College of Physicians and Surgeons, 630 West 168th Street, New York, New York 10032, USA. [2]Department of Psychiatry, Columbia University College of Physicians and Surgeons, 630 West 168th Street, New York, New York 10032, USA. [3]Department of Pharmacology, Columbia University College of Physicians and Surgeons, 630 West 168th Street, New York, New York 10032, USA. [4]Division of Molecular Therapeutics, New York State Psychiatric Institute, 1051 Riverside Drive, New York, New York 10032, USA. [5]Department of Physiology and Biophysics, Weill Medical College of Cornell University, 1300 York Avenue, New York, New York 10021, USA. [6]HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medical College, Cornell University, 1300 York Avenue, New York, New York 10021, USA.
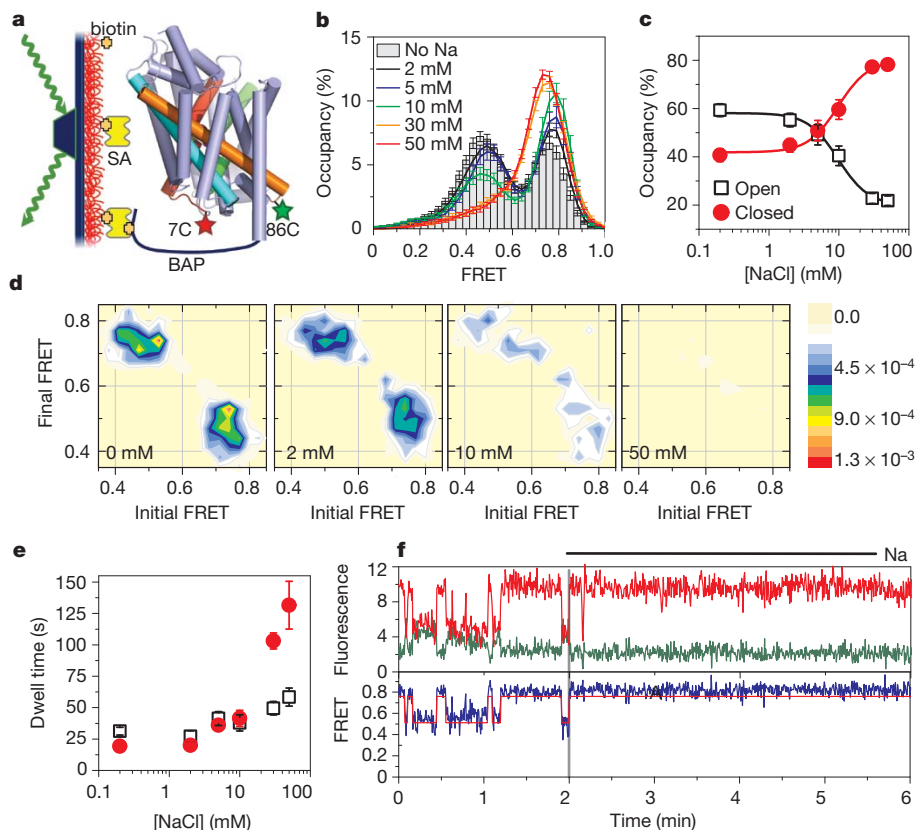*These authors contributed equally to this work.

Figure 1 | Effect of Na+ on LeuT dynamics. a, Experimental set-up: H7C/R86C-LeuT labelled with Cy3 and Cy5 (stars) was immobilized via a biotin acceptor peptide (BAP) on a passivated glass surface and illuminated using total internal reflection. FRET traces (>110 per condition) were collected with varying concentrations of Na+ (160-ms time resolution for all, except 30–50 mM with 400 ms). **b**, Histograms of FRET traces, filtered to remove fluorophore dark states. **c**, Fraction of time in the lower-FRET open state (black open squares) and the high-FRET closed state (red filled circles). **d**, Transition density plot: average FRET values before (x axis) and after (y axis) each transition were plotted as a two-dimensional chart in transitions per second (scale at right; Na+ concentrations are indicated). **e**, Average dwell times in each state. **f**, Representative traces (donor in green, acceptor in red, FRET in blue, and predicted state sequence (idealization) in red), where the solution was exchanged at 2 min from K+ to Na+ (200 mM). Error bars, s.d. of ≥100 bootstrap samples.

observed for H7C/T515C-LeuT (Supplementary Fig. 3). In accordance with such effects, which required both Na+ and Ala, the lifetimes of the inward-open or inward-closed FRET states were not significantly affected by Ala alone (in the nominal absence of Na+); at 250 μM Ala, the transition frequency increased in a Na+-concentration-dependent fashion (Supplementary Fig. 4).

Using transition state theory (Methods), we found that the intracellular open and closed FRET states of LeuT were separated by a large activation barrier ($\Delta G^{\ddagger} \approx 80 \text{ kJ mol}^{-1}$). Ala does not alter the relative occupancies of open and closed states, but instead lowers the activation barrier for both open-to-closed and closed-to-open transitions by approximately 3 kJ mol^{-1} (about the energy of a hydrogen bond). By contrast, Leu raised the activation barrier for the closed-to-open transition by as much as 4 kJ mol^{-1}, apparently through ground-state stabilization of the closed state.

Hypothesizing that the observed dynamics reflect Ala's acceleration of the opening-closing cycles of the intracellular gate required for the transport mechanism, we performed experiments in the presence of the transport inhibitor clomipramine (CMI), a tricyclic antidepressant that is known to bind in an extracellular vestibule above the Na+ and S1 binding sites[25,27,28]. Many of the residues shown to interact with antidepressants in these structures are also part of the S2 site[25,27]. As substrate binding in the S2 site is thought to allosterically trigger intracellular release of Na+ and substrate from the S1 site[15] (also see Supplementary Fig. 1), CMI should block Ala-induced intracellular gating dynamics. Indeed, in the presence of both Na+ and Ala, CMI essentially eliminated intracellular gate opening, stabilizing LeuT in a high-FRET, inward-closed conformation (Supplementary Fig. 5a–c).

This observation is consistent with CMI competitively blocking substrate binding to the S2 site[15], thereby preventing Ala-induced opening and closing of the intracellular gate, and inhibiting transport. This result was again confirmed by direct imaging of individual LeuT molecules in Na+ and Ala-containing buffer on addition of CMI (Fig. 2f). The detergent n-octyl-β-D-glucopyranoside also inhibited intracellular gating dynamics (Supplementary Fig. 5a–c), consistent with its capacity to disrupt the Na+-coupled transport mechanism[20] by competing with substrate binding to the S2 site[20,23].

To probe whether Ala binding to the S1 and/or S2 site(s) was responsible for lowering the activation barrier for intracellular gating dynamics, single-molecule FRET experiments were performed in the background of either an F253A or L400S mutation within the S1 or S2 site, respectively (Fig. 3a, Supplementary Fig. 1). These mutations disrupt substrate binding to LeuT, decreasing the stoichiometry of substrate:LeuT binding under saturating conditions from 2:1 in wild-type LeuT, to 1:1 in both mutants (Fig. 4a). Mutation of F253 blocks substrate binding to the S1 site and also abrogates transport (Fig. 4a, b; Supplementary Fig. 6), while having little or no effect on Na+ binding (Supplementary Table 1). Despite evidence that Ala bound to the S2 site in the context of the F253A mutation (Fig. 4a), Ala failed to increase intracellular gating dynamics of the mutant protein (Fig. 4c). Similarly, despite evidence of Ala binding to the S1 site (Fig. 4a), no increase in intracellular gating dynamics was observed when the S2 site was disrupted by the L400S mutation (Fig. 4c). These findings support the notion that substrate occupancy in the S2 site is critical for the allosteric mechanism that controls intracellular gate opening and the release of substrate from the S1
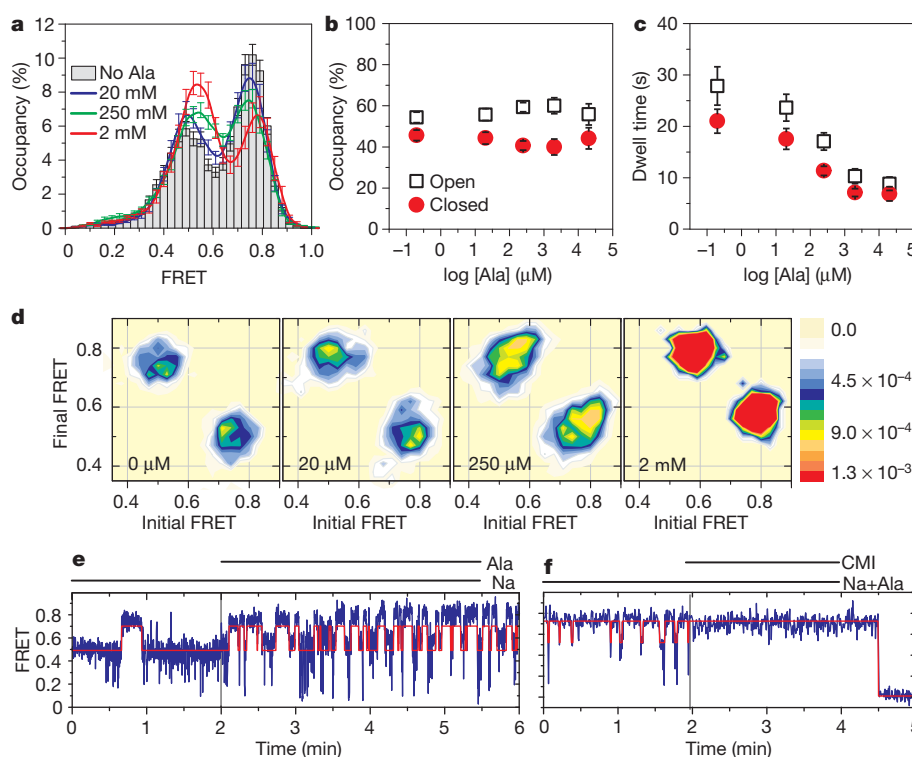
Figure 2 | Effect of alanine on LeuT dynamics. Single-molecule FRET traces (>90 per condition) were collected at 160-ms time resolution with 2 mM Na$^+$ and varying concentrations of Ala. **a**, Histograms of FRET data from each condition. Hidden Markov Modelling analysis revealed the fraction of time (**b**) and average dwell times (**c**) in the lower-FRET open state (black open squares) and the high-FRET closed state (red filled circles). **d**, Transition density plots as in Fig. 1d (Ala concentrations are indicated). **e**, **f**, Representative FRET traces (blue) with idealization (red) from experiments where solution was exchanged at 2 min: **e**, 2 mM Na$^+$, adding 250 μM Ala; **f**, 2 mM Na$^+$ and 250 μM Ala, adding the inhibitor clomipramine (CMI; 0.5 mM). Error bars, s.d. of ≥100 bootstrap samples.

site[15], and demonstrate that substrate binding to both the S1 and S2 sites is necessary to trigger intracellular gating.

In order to probe whether Ala binding to the S1 and S2 sites is also sufficient to promote intracellular gating and transport, experiments were performed in the presence of Li$^+$ in place of Na$^+$. In the presence of saturating Li$^+$ concentrations (>150 mM), we found that Ala binds LeuT with a 2:1 stoichiometry consistent with both S1 and S2 site occupancy (Fig. 4a). Li$^+$, like Na$^+$, stabilized the inward-closed state (Supplementary Fig. 7), but, in the presence of Li$^+$, Ala failed to accelerate intracellular gating dynamics and no substrate transport was

observed (Fig. 4c). Instead, the inward-closed conformation of LeuT was modestly stabilized in the presence of Ala (~2-fold reduction in the rate of gate opening, $k_{closed-open}$) (Fig. 4c). These data demonstrate that Ala binding to the S1 and S2 sites in the presence of Li$^+$ does not lower the activation barrier to intracellular gating as observed in the presence of Na$^+$.

Prompted by these experimental observations, computational studies were performed to investigate how both Na$^+$ and Li$^+$ can support substrate binding to LeuT, whereas only Na$^+$ leads to substrate-induced dynamics of the intracellular gate and to transport. These studies also
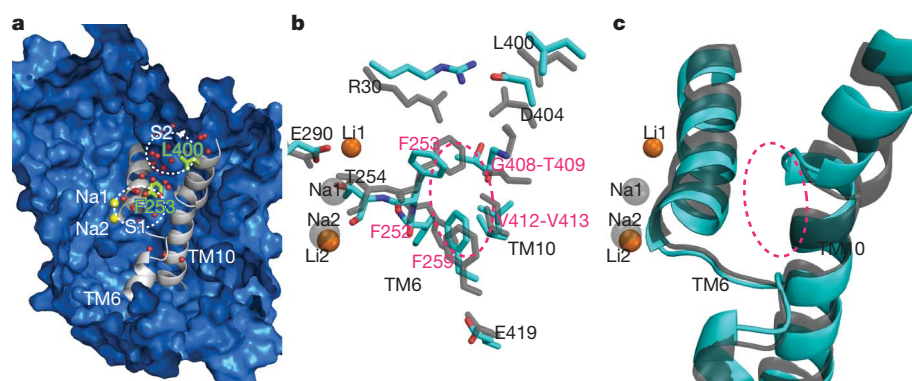

Figure 3 | The configuration of TM6–TM10 interactions induced by Na$^+$ binding cannot be matched by Li$^+$ binding. a, Representative snapshot taken from the Na-only simulation, showing water molecules (red spheres) occupying the S1 and S2 sites (white dotted ellipses). Residues L400 in the S2 site and F253 in the S1 site, which were mutated to affect substrate binding, are shown as light green sticks. **b**, The different effects that Li$^+$ and Na$^+$ binding

have on the interacting residues of TM6 and TM10. The TM6/TM10 interface is indicated by the dashed ellipse in magenta. **c**, The bulge around G408 in TM10, which is present only when Na$^+$ is bound but not when Li$^+$ replaces it. In **b** and **c**, side chains and backbones coloured according to atom types are from the Li-only conformation, while those from the Na-only conformation are rendered in grey.
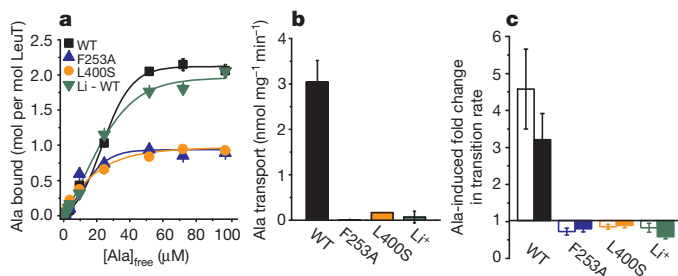
Figure 4 | Effect of S1 and S2 site mutations and of Li^+ on activity and dynamics. a, Binding of ^3H-Ala in buffer containing 50 mM Na^+ was measured for wild-type (WT, black squares), F253A (blue triangles) and L400S (orange circles) LeuT, and for wild-type LeuT with 150 mM Li^+ (green inverted triangles). **b**, ^3H-Ala uptake with 100 mM Na^+ was measured for WT (black), F253A (blue) and L400S (orange) LeuT or with 150 mM Li^+ for WT LeuT (green). Error bars in **a** and **b**, s.e.m. of triplicate determinations. **c**, The fold change in the rate of transitioning from the open state to the closed state (open bars) and from the closed state to the open state (filled bars) induced by 250 μM Ala in 10 mM Na^+ or for WT in 40 mM Li^+ (>280 traces and 800 transitions per condition). Error bars in **c**, s.d. of ≥100 bootstrap samples.

served to identify both local changes produced in the region of the ion binding sites and critical elements in the allosteric pathway linking the substrate binding sites and the intracellular gate region. Comparative analysis of separate molecular dynamics (MD) simulations of LeuT, performed with either Na^+ or Li^+ occupying the established Na^+ binding sites and in the absence of amino acid substrate (termed Na-only[24] and Li-only, respectively) revealed significant differences in TM–TM interactions (Fig. 3b, c), which are described in detail in Supplementary Information. The Na1/Li1 binding site and its neighbouring interaction network, which are crucial for the proper propagation of the allosteric effects from the S2 to S1 site (see Supplementary Information for details) and onward to the intracellular side to open the transport pathway, are sensitive to the unique combination of the ionic radius of the Na^+ cation and the charge redistribution it causes. The structural consequences of the ion-specific effects appear to be propagated through the cluster of aromatic residues at the heart of the S1 binding site, and result in different configurations of the bulge in the middle of TM10 (Fig. 3, Supplementary Fig. 8).

The positions of the structural elements involved in this propagation mechanism make them critical for transmitting conformational changes deeper into the TM bundle towards the intracellular end of the transporter (Fig. 3b, c). Such changes include local alterations in the vicinity of E419, a residue known from the crystal structure to interact with E62 in TM2, with the backbone of the unwound portion of TM6 (proximal to F259 of the aromatic cluster and the S1 binding site), and two water molecules[29]. Reconfiguration of this region, including residue T418, on simulated inward movement of the substrate[15] was previously shown to enable the penetration of water from the intracellular side of LeuT as a result of an opening at IL1[15]. The resulting dissociation of IL1 from interactions with R5 and D369 and the destabilization of the network of intracellular interactions detected in the simulations (Supplementary Fig. 9) is associated with the observed outward movement of TM1a[22] that is essential for the simulated release of substrate to the intracellular side.

Owing to the different effects of Li^+ and Na^+, Ala binding in both the S1 and S2 sites in the presence of Li^+ would not engender the ordered series of local conformational rearrangements expected in the presence of Na^+. These rearrangements originate in the S2 site and need to be propagated as described above through changes in the Na1 and S1 sites to enable water penetration from the cytoplasmic side of LeuT and the outward movement of TM1a. Their absence when Li^+ substitutes for Na^+ would explain why substrate-induced acceleration of gating dynamics was not observed experimentally.

Na^+ binding, which stabilizes the inward-closed state, does not hasten gate closure but, instead, slightly stabilizes the inward-open

state as well, by raising the energy barrier to the conformational transition. In contrast, Ala binding to LeuT in the presence of Na^+ shortens not only the inward-closed, but also the inward-open, lifetime (Fig. 2). Thus, bound Ala facilitates both the opening of the intracellular gate and its subsequent closure by reducing the activation barrier for such conformational transitions. One possible explanation for this observation is that binding of substrate in the S2 site triggers the opening of the intracellular gate and release of the S1 substrate to the cytoplasm. In the absence of S1 substrate and bound Na^+, substrate in the S2 site may then facilitate intracellular gate closure. It is tempting to speculate that the S2 substrate, in the presence of extracellular Na^+, may move to the S1 site with high efficiency owing to its very high local concentration, thereby facilitating a subsequent transport cycle.

Collectively, our findings support the notion that the observed movements of TM1a and its environment are associated with LeuT intracellular gating[22] in a manner that is directly linked to the Na^+-driven transport mechanism. Thus, results obtained with the slowly transported substrate, Leu, and the relatively rapidly transported substrate, Ala, establish a relationship between the rates of intracellular gating and substrate transport. The role of substrate binding at the S2 site in the process of allostery and molecular recognition is further highlighted by the comparative effects of CMI and Ala binding to this site in the presence of Na^+. The former stabilizes a closed intracellular gate conformation, whereas the latter substantially lowers the activation barrier to gate opening and thereby allows the energy of the Na^+ gradient to drive the transport mechanism.

After the present manuscript had been submitted, a report[30] was published that concluded, on the basis of a variety of binding measurements, that LeuT has only a single high-affinity substrate site. In contrast, our substrate binding measurements clearly show a stoichiometry of 2:1, consistent with high affinity binding to both the S1 and S2 sites[20]. Half of this binding is lost in the S2-site mutant that also exhibits a loss of substrate-induced single-molecule dynamics and transport (Fig. 4). Although the loss of substrate-induced dynamics and transport in the S2-site mutant could conceivably be explained solely by a long-range allosteric effect of the mutation, all our data to date are most consistent with a two-substrate-site model in which the absence of either S1 or S2 substrate binding results in a profound attenuation of transporter dynamics and function. We are currently uncovering the reasons for the discrepancy between the data of ref. 30 and our own data, and will report our findings in due course.

METHODS SUMMARY

LeuT mutants were expressed in *Escherichia coli*, purified, and labelled on targeted engineered cysteines with Cy3 and Cy5 maleimide. The functional properties of the labelled constructs were determined by measuring Leu binding and Na^+ by scintillation proximity assay, and Ala transport was measured after reconstitution of the protein into proteoliposomes. Purified, labelled protein was immobilized onto a passivated glass surface via a streptavidin-biotin linkage (shown schematically in Fig. 1a). Fluorescence data were acquired using a prism-based total internal reflection (TIR) microscope. FRET efficiency was calculated and analysis of fluorescence and FRET traces was achieved using semi-automated analysis software developed for this application. The single-molecule traces were analysed for LeuT in the presence and absence of the substrates Na^+, Leu and Ala, and on addition of the transport inhibitors CMI and n-octyl-β-D-glucopyranoside, and in response to mutations of the S1 and S2 binding sites. MD simulations were carried out with the protein immersed in an all-atom model of the membrane, solvated with water molecules, ions and ligands. Long equilibrations (totalling >2 μs) were run to assess conformational changes, with more than one MD trajectory collected for every configuration mentioned.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

1. Amara, S. G. & Sonders, M. S. Neurotransmitter transporters as molecular targets for addictive drugs. *Drug Alcohol Depend.* **51,** 87–96 (1998).

2. Rudnick, G. *Mechanisms of Biogenic Amine Neurotransmitter Transporters* 2nd edn (Humana, 2002).
3. Sonders, M. S., Quick, M. & Javitch, J. A. How did the neurotransmitter cross the bilayer? A closer view. *Curr. Opin. Neurobiol.* **15,** 296–304 (2005).
4. Gu, H., Wall, S. C. & Rudnick, G. Stable expression of biogenic amine transporters reveals differences in inhibitor sensitivity, kinetics, and ion dependence. *J. Biol. Chem.* **269,** 7124–7130 (1994).
5. Torres, G. E., Gainetdinov, R. R. & Caron, M. G. Plasma membrane monoamine transporters: structure, regulation and function. *Nature Rev. Neurosci.* **4,** 13–25 (2003).
6. Krause, S. & Schwarz, W. Identification and selective inhibition of the channel mode of the neuronal GABA transporter 1. *Mol. Pharmacol.* **68,** 1728–1735 (2005).
7. Iversen, L. Neurotransmitter transporters and their impact on the development of psychopharmacology. *Br. J. Pharmacol.* **147** (Suppl 1), S82–S88 (2006).
8. Yamashita, A. *et al.* Crystal structure of a bacterial homologue of Na^+/Cl^--dependent neurotransmitter transporters. *Nature* **437,** 215–223 (2005).
9. Beuming, T., Shi, L., Javitch, J. A. & Weinstein, H. A comprehensive structure-based alignment of prokaryotic and eukaryotic neurotransmitter/Na^+ symporters (NSS) aids in the use of the LeuT structure to probe NSS structure and function. *Mol. Pharmacol.* **70,** 1630–1642 (2006).
10. Quick, M. *et al.* State-dependent conformations of the translocation pathway in the tyrosine transporter Tyt1, a novel neurotransmitter:sodium symporter from *Fusobacterium nucleatum. J. Biol. Chem.* **281,** 26444–26454 (2006).
11. Forrest, L. R. *et al.* Mechanism for alternating access in neurotransmitter transporters. *Proc. Natl Acad. Sci. USA* **105,** 10338–10343 (2008).
12. Kniazeff, J. *et al.* An intracellular interaction network regulates conformational transitions in the dopamine transporter. *J. Biol. Chem.* **283,** 17691–17701 (2008).
13. Noskov, S. Y. Molecular mechanism of substrate specificity in the bacterial neutral amino acid transporter LeuT. *Proteins* **73,** 851–861 (2008).
14. Noskov, S. Y. & Roux, B. Control of ion selectivity in LeuT: two Na^+ binding sites with two different mechanisms. *J. Mol. Biol.* **377,** 804–818 (2008).
15. Shi, L. *et al.* The mechanism of a neurotransmitter:sodium symporter — inward release of Na^+ and substrate is triggered by substrate in a second binding site. *Mol. Cell* **30,** 667–677 (2008).
16. Singh, S. K. LeuT: a prokaryotic stepping stone on the way to a eukaryotic neurotransmitter transporter structure. *Channels (Austin)* **2,** 380–389 (2008).
17. Crisman, T. J., Qu, S., Kanner, B. I. & Forrest, L. R. Inward-facing conformation of glutamate transporters as revealed by their inverted-topology structural repeats. *Proc. Natl Acad. Sci. USA* **106,** 20752–20757 (2009).
18. Khalili-Araghi, F. *et al.* Molecular dynamics simulations of membrane channels and transporters. *Curr. Opin. Struct. Biol.* **19,** 128–137 (2009).
19. Li, J. & Tajkhorshid, E. Ion-releasing state of a secondary membrane transporter. *Biophys. J.* **97,** L29–L31 (2009).
20. Quick, M. *et al.* Binding of an octylglucoside detergent molecule in the second substrate (S2) site of LeuT establishes an inhibitor-bound conformation. *Proc. Natl Acad. Sci. USA* **106,** 5563–5568 (2009).
21. Shi, L. & Weinstein, H. Conformational rearrangements to the intracellular open states of the LeuT and ApcT transporters are modulated by common mechanisms. *Biophys. J.* **99,** L103–L105 (2010).
22. Zhao, Y. *et al.* Single-molecule dynamics of gating in a neurotransmitter transporter homologue. *Nature* **465,** 188–193 (2010).
23. Singh, S. K., Piscitelli, C. L., Yamashita, A. & Gouaux, E. A competitive inhibitor traps LeuT in an open-to-out conformation. *Science* **322,** 1655–1661 (2008).
24. Claxton, D. P. *et al.* Ion/substrate-dependent conformational dynamics of a bacterial homolog of neurotransmitter:sodium symporters. *Nature Struct. Mol. Biol.* **17,** 822–829 (2010).
25. Singh, S. K., Yamashita, A. & Gouaux, E. Antidepressant binding site in a bacterial homologue of neurotransmitter transporters. *Nature* **448,** 952–956 (2007).
26. Beckett, D., Kovaleva, E. & Schatz, P. J. A minimal peptide substrate in biotin holoenzyme synthetase-catalyzed biotinylation. *Protein Sci.* **8,** 921–929 (1999).
27. Zhou, Z. *et al.* LeuT-desipramine structure reveals how antidepressants block neurotransmitter reuptake. *Science* **317,** 1390–1393 (2007).
28. Zhou, Z. *et al.* Antidepressant specificity of serotonin transporter suggested by three LeuT-SSRI structures. *Nature Struct. Mol. Biol.* **16,** 652–657 (2009).
29. Sen, N., Shi, L., Beuming, T., Weinstein, H. & Javitch, J. A. A pincer-like configuration of TM2 in the human dopamine transporter is responsible for indirect effects on cocaine binding. *Neuropharmacology* **49,** 780–790 (2005).
30. Piscitelli, C. L., Krishnamurthy, H. & Gouaux, E. Neurotransmitter/sodium symporter orthologue LeuT has a single high-affinity substrate site. *Nature* **468,** 1129–1132 (2010).

Author Contributions Y.Z. expressed, purified and labelled the LeuT mutants. M.Q. and Y.Z. performed the functional characterization of the mutants. Y.Z. and D.S.T. designed, carried out and analysed the single-molecule experiments; L.S. and H.W. designed and analysed the computational studies, which were carried out by L.S.; S.C.B. and J.A.J. helped to design the biochemical and single-molecule experiments and, with L.S. and H.W., helped to interpret the data. All the authors contributed to writing and editing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.A.J. (jaj2@columbia.edu) or S.C.B (scb2005@med.cornell.edu).

METHODS

Protein expression and purification. LeuT variants were expressed in *E. coli* C41(DE3) as described[15]. For functional studies, LeuT variants were expressed from pQO18 or derivatives thereof carrying the indicated mutations[20], whereas for single-molecule FRET studies, biotin acceptor peptide-tagged LeuT variants were expressed in pETO18G and its derivatives[22]. Protein was purified by immobilized metal (Ni^{2+}) affinity chromatography using a Ni^{2+} Sepharose 6 FastFlow column (GE Healthcare)[22]. For fluorescent labelling of LeuT, Cy3-maleimide and Cy5-maleimide (GE Healthcare) were added at an equimolar ratio (200 μM total) for 1 h while the protein was bound to the Ni^{2+} resin[22]. Free dye was removed before the elution of LeuT with 300 mM imidazole.

Scintillation proximity-based binding studies. Binding of ^{3}H-leucine or ^{3}H-alanine (146 Ci mmol^{-1} and 49.4 Ci mmol^{-1}, respectively; both from Moravek) to purified LeuT-variants was measured with the scintillation proximity assay (SPA) as described[15], with 25 ng of purified protein per assay in buffer composed of 150 mM Tris/MES, pH 7.5/50 mM NaCl/1 mM TCEP/0.1% n-dodecyl-β-D-maltopyranoside or 50 mM Tris/MES, pH 7.5/150 mM LiCl/1 mM TCEP/0.1% n-dodecyl-β-D-maltopyranoside. To determine the molar ratio of Leu (or Ala)-to-LeuT, binding samples were incubated with increasing concentrations of radioligand and measured in the SPA c.p.m. mode of the Wallac 1450 MicroBeta counter (Perkin Elmer). The efficiency of detection was calculated with standard curves of known concentrations of ^{3}H-Leu or ^{3}H-Ala. The standard curves were used to transform c.p.m. into the amount of bound substrate[15]. The amount of LeuT in the SPA assays was determined[31]. SPA-based binding studies using 2 μM [^{22}Na]Cl (1,017 mCi mg^{-1}; Perkin Elmer) were performed in 150–200 mM Tris/MES, pH 7.5/1 mM TCEP/0.1% n-dodecyl-β-D-maltopyranoside in the presence of 0–50 mM NaCl (equimolar replacement of Tris/MES to maintain a total molarity of 200 mM)[15]. All experiments were repeated at least in duplicate with triplicate determination of all individual data points. Kinetic constants (shown ± the s.e.m. of the fit) were obtained by fitting the data from independent experiments to global fitting in Prism or SigmaPlot.

^{3}H-Ala transport in proteoliposomes. Proteoliposomes were prepared as described[15]. The accumulation of ^{3}H-Ala (49.4 Ci mmol^{-1}; Moravek) was measured at 23 °C in assay buffer composed of 150/50 mM Tris/MES (pH 8.5) and 50 mM NaCl/150 mM LiCl. The reaction was quenched by the addition of ice-cold assay buffer without radiotracer and the proteoliposomes were collected on GF-75 glass fibre filters (Advantec) before the determination of the accumulated c.p.m. by liquid scintillation counting.

Single-molecule FRET imaging experiments. Fluorescence experiments were performed using a prism-based total internal reflection fluorescence (TIRF) microscope as previously described[22,32]. Microfluidic imaging chambers were passivated with a mixture of PEG and biotin-PEG and incubated with 0.8 μM streptavidin (Invitrogen). Cy3/Cy5-labelled, biotinylated LeuT molecules were surface immobilized through biotin-streptavidin interaction. Cy3 fluorophores were excited by the evanescent wave generated by total internal reflection (TIR) of a single-frequency light source (Ventus 532, Laser Quanta). Photons emitted from Cy3 and Cy5 were collected using a 1.2 NA 60× water-immersion objective (Nikon) and optical treatments were used to separate Cy3 and Cy5 frequencies onto a cooled, back-thinned EMCCD camera (Cascade 512, Photometrics). Fluorescence data were acquired using Metamorph (Universal Imaging Corporation).

All experiments were performed in buffer containing 50 mM Tris/MES, pH 7.5, 10% glycerol, 0.02% (w/v) DDM, 5 mM 2-mercaptoethanol and 200 mM salt (KCl or NaCl, as specified). We used an oxygen-scavenging environment (1 unit per ml glucose oxidase, 8 units per ml catalase, 0.1% (v/v) glucose) containing 2 mM cyclooctatetraene in all experiments to minimize photobleaching.

Analysis of single-molecule fluorescence data was performed using custom software written in MATLAB (MathWorks). A subset of the acquired traces was selected for further analysis using the following criteria: (1) single-step donor photobleaching, (2) signal-to-background noise ratio (SNR) ≥ 8, (3) <4 donor blinking events, (4) non-zero FRET efficiency for at least 60 s. Additional manual trace selection was performed to refine the data, where selected traces were required to have: (1) stable total fluorescence intensity ($I_D + I_A$) and (2) at least one transition between clearly defined FRET states with anti-correlated transitions in donor/acceptor intensity or a single dwell in a clearly-defined FRET state. We found this process to be effective in removing artefacts and spurious noise without introducing significant bias (see Supplementary Discussion and Supplementary Fig. 10).

Kinetic analysis was performed to idealize FRET traces and calculate average dwell times using a three-state model as previously described[22]. Error bars for transition rate estimates and FRET histograms were calculated as the standard deviation of 100 bootstrap samples of the traces. Error bars for state occupancies were calculated from 1,000 bootstrap samples.

Transition rates were interpreted using transition state theory, where the open and closed states are considered ground states separated by a large ($\Delta G^{\ddagger} \approx 80$ kJ mol^{-1}) activation barrier (the transition state). The energy required to achieve the transition state (and cross the barrier) was calculated as:

$$\Delta G^{\ddagger} = -RT \ln\left(\frac{hk_{i,j}}{k_B T}\right),$$

where R is the gas constant, T is absolute temperature (296 K), h is Planck's constant, k is the measured transition rate, from state i to state j, and k_B is Boltzmann's constant. Changes in the activation barrier energy ($\Delta\Delta G^{\ddagger}$) were calculated from the difference in forward and reverse rates observed in the absence and presence of substrate.

Molecular dynamics. The Li$^+$-only simulation was performed on a system prepared as described[24]. Briefly, it consisted of more than 77,000 atoms, including the explicit membrane model, solvating water molecules and the various ions and ligands. All the Na$^+$ ions in the system were replaced with Li$^+$. The parameters for Li$^+$ were from ref. 33. All MD simulations were carried out with the NAMD program under constant temperature (310 K) and constant pressure (1 atm) conditions. Long equilibration runs were performed to allow the system to transition to a new stable conformation. The inward-closed and inward-open conformations described in Supplementary Fig. 9 were based on the simulations described previously[22]. More than one MD trajectory was collected for every configuration studied. Each individual trajectory was at least 360 ns, and the longest trajectory for each configuration was 720 ns. All the results reinforced the conclusions, and the structural and dynamic insights described in the main text were revealed as the common features and trends of parallel independent MD runs.

31. Schaffner, W. & Weissmann, C. A rapid, sensitive, and specific method for the determination of protein in dilute solution. *Anal. Biochem.* **56,** 502–514 (1973).
32. Munro, J. B., Altman, R. B., O'Connor, N. & Blanchard, S. C. Identification of two distinct hybrid state intermediates on the ribosome. *Mol. Cell* **25,** 505–517 (2007).
33. Caplan, D. A., Subbotina, J. O. & Noskov, S. Y. Molecular mechanism of ion-ion and ion-substrate coupling in the Na$^+$-dependent leucine transporter LeuT. *Biophys. J.* **95,** 4613–4621 (2008).

Royalactin induces queen differentiation in honeybees

Masaki Kamakura[1]

The honeybee (*Apis mellifera*) forms two female castes: the queen and the worker. This dimorphism depends not on genetic differences, but on ingestion of royal jelly, although the mechanism through which royal jelly regulates caste differentiation has long remained unknown. Here I show that a 57-kDa protein in royal jelly, previously designated as royalactin, induces the differentiation of honeybee larvae into queens. Royalactin increased body size and ovary development and shortened developmental time in honeybees. Surprisingly, it also showed similar effects in the fruitfly (*Drosophila melanogaster*). Mechanistic studies revealed that royalactin activated p70 S6 kinase, which was responsible for the increase of body size, increased the activity of mitogen–activated protein kinase, which was involved in the decreased developmental time, and increased the titre of juvenile hormone, an essential hormone for ovary development. Knockdown of epidermal growth factor receptor (Egfr) expression in the fat body of honeybees and fruitflies resulted in a defect of all phenotypes induced by royalactin, showing that Egfr mediates these actions. These findings indicate that a specific factor in royal jelly, royalactin, drives queen development through an Egfr–mediated signalling pathway.

Caste in social insects represents one of the major transitions from one level of organization to another in evolution[1]. The honeybee (*Apis mellifera*) exhibits polyphenism, that is, adult females form two interdependent castes, the queen and the worker, depending on their environment at critical periods of caste determination[2,3]. This dimorphism is not a consequence of genetic difference[4,5]. Queens have a larger body size and shorter developmental time than workers, have ten times the lifespan of workers, typically 1 to 2 years, and lay up 2,000 eggs per day, whereas workers rear young larvae and gather nectar[6,7]. When larvae are nourished with royal jelly, which is secreted by workers[2,3], they differentiate into queens. Royal jelly seems to contain a specific factor(s) that determines caste differentiation, but this has not previously been identified. Furthermore, the relationship between caste-specific modulation of juvenile hormone and ecdysteroid after ingestion of royal jelly and the developmental signal in caste differentiation has remained elusive. Therefore, I aimed to identify the factor(s) that induces caste differentiation in the honeybee and to investigate the mechanism through which this factor drives the caste-specific developmental pathway.

A caste differentiation–inducing factor in royal jelly

The dietary requirements for rearing queens are known[8], but a diet for rearing worker honeybees has not been reported. In connection with this, I found that larvae reared with royal jelly stored at 40 °C for 7 days, which did not exhibit any antifatigue effect[9], showed increased developmental times, decreased body weight at eclosion and decreased ovary size, compared to larvae fed a diet containing fresh royal jelly, even though they were queen–worker intermediates (Supplementary Fig. 1a–c). This result indicated that long-term storage of royal jelly at high temperature decreases the biological activity of royal jelly for queen differentiation. Therefore, royal jelly was stored at 40 °C for 7, 14, 21 and 30 days, and the effects of these royal jelly samples on caste differentiation were examined. Storage of royal jelly at 40 °C for up to 30 days caused a reduction in the growth of developing larvae, decreased weight at adult emergence, ovary size reduction and prolongation of

the pre-adult development time in proportion to storage duration (Supplementary Fig. 1). Adult females reared with royal jelly stored at 40 °C for 30 days (40 °C/30 d royal jelly) developed with a full worker morphotype. These results indicate that the putative inducer of queen differentiation in royal jelly might be gradually degraded in proportion to the storage period at 40 °C, being completely degraded after 30 days. Therefore, the compositional changes in royal jelly during storage were investigated next.

First, the contents of several vitamins, 10-hydroxy-2-decenoic acid, carbohydrates and fatty acids in royal jelly samples stored at 4 °C and 40 °C for 30 days were measured. No significant differences were observed in the contents of the examined compounds, except pantothenic acid, which showed a decrease to 60% of the initial concentration during storage at 40 °C for 30 days (Supplementary Table 1). However, pantothenic acid did not induce the emergence of queens (data not shown), in agreement with a previous report[10]. Next, compositional changes of proteins in royal jelly during storage were analysed by means of high-performance liquid chromatography (HPLC) and native polyacrylamide gel electrophoresis (PAGE). A 450-kDa protein, a 170-kDa protein and a 57-kDa protein (designated as royalactin[11]) were degraded during storage (Supplementary Fig. 2a and Supplementary Fig. 3). Royalactin is a monomeric protein that exhibits epidermal growth factor (Egf)-like effects on rat hepatocytes[11,12]. The 170-kDa protein was completely degraded during storage at 40 °C for 14 days, being undetectable in royal jelly stored at 40 °C for 21 or 30 days; because this royal jelly can still influence ovary development and growth of developing larvae, the 170-kDa protein seems to be irrelevant to caste differentiation (Supplementary Figs 1 and 2a). Royalactin was degraded proportionally to the period of storage, and was completely lost during storage at 40 °C for 30 days, whereas only 10% of the 450-kDa protein was destroyed during storage at 40 °C for 30 days (Supplementary Fig. 2a).

Next, royalactin and the 450-kDa protein were purified (Supplementary Fig. 2b–d), and the effects of these factors on caste differentiation were examined in the same manner described above. As

---

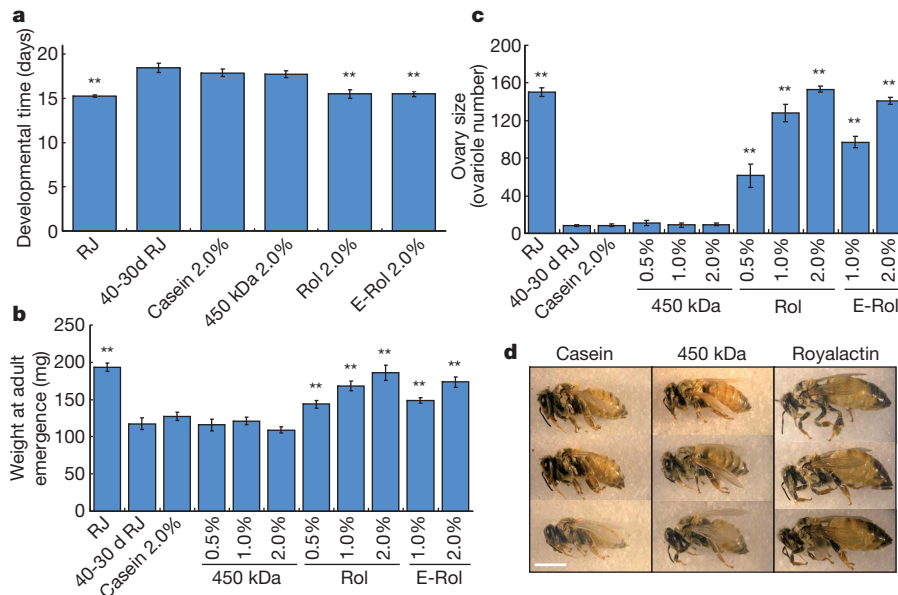[1]Biotechnology Research Center, Toyama Prefectural University, Imizu, Toyama 939-0398, Japan.

Figure 1 | Effects of casein, 450-kDa protein, royalactin and recombinant royalactin on caste characters in the honeybee. a–c, Developmental time (**a**), weight at adult emergence (**b**) and ovary size (**c**) in individuals (*n* = 10–28) reared with royal jelly, royal jelly stored at 40 °C for 30 days (40-30d RJ) or 40- 30d RJ containing casein, 450-kDa protein, royalactin (Rol) or E-royalactin (E-Rol) were measured. **d**, Final adult size after eclosion is shown. Values are expressed as mean ± s.e.m. Values significantly different from those of larvae reared with 40-30d RJ are indicated by ****P < 0.01. Scale bar, 5 mm.

shown in Fig. 1, the 450-kDa protein (0.5% to 2.0% w/w diet) and casein (2.0% w/w diet), which was used as a control for evaluating nutritional effect, did not change the final adult size, developmental time, or ovary size in individuals reared with 40 °C/30 d royal jelly. In contrast, royalactin shortened developmental time and increased both weight at adult emergence and ovary size in proportion to the concentration added to a diet containing 40 °C/30 d royal jelly, and it induced larvae to develop into queens as effectively as did royal jelly at the concentration of 2.0% w/w diet (Fig. 1 and Supplementary Fig. 4a). Similar results were observed in larvae reared with recombinant royalactin (E-royalactin; 47 kDa), which was expressed in *Escherichia coli* and purified to homogeneity on SDS–PAGE (Fig. 1, Supplementary Fig. 2e and Supplementary Fig. 4a). Furthermore, royalactin and E-royalactin increased the juvenile hormone titre—which increases at the fourth larval instar to cause development into a queen[13,14]—in larvae given 40 °C/30 d royal jelly as potently as royal jelly, whereas the 450-kDa protein or casein had no effect (Supplementary Fig. 4b). Taken together, these results indicate that the stimulatory effect of royalactin on caste differentiation was not a nutritional effect but a morphogenic effect, and that royalactin is the major active factor in the induction of caste differentiation by royal jelly.

Effects of royal jelly and royalactin on *Drosophila*

Because no mutant stock of *Apis mellifera* has so far been developed, it is difficult to investigate the mechanism underlying honeybee caste differentiation at the individual level. On the other hand, fruitfly (*Drosophila melanogaster*), used as a model organism in many research fields, is available for genetic analysis in developmental biology. I considered that *Drosophila* might be suitable as a model insect for analysis of the mechanism of caste differentiation if royal jelly induced morphological and physiological changes in *Drosophila* similar to those induced in honeybee queens. Therefore, I investigated the influence of royal jelly on *Drosophila* larvae.

When *Drosophila* (Canton-S) larvae were reared with only royal jelly, they died before pupation (data not shown). However, *Drosophila* reared with medium containing 20% royal jelly, 8% yeast and 10% D-glucose had an increase in body size (body weight and body length) and fecundity, and had extended lifespan and shortened developmental time compared to flies reared with control medium or casein

medium, which provide the same total energy as royal jelly medium (Fig. 2 and Supplementary Table 2). Furthermore, royal jelly medium increased cell size but not cell number (Supplementary Fig. 5). Royalactin increased body size, cell size and fecundity, extended lifespan and shortened developmental time in flies reared with 40 °C/30 d royal jelly (which did not influence morphological or physiological changes of flies), whereas 450-kDa protein or casein did not (Fig. 2, Supplementary Fig. 5 and Supplementary Table 2), in accordance with the observations that royalactin induced queen differentiation in honeybee as the major active factor in royal jelly. Thus, fresh royal jelly led genetically identical fly larvae to develop into adult individuals with phenotypes similar to queen bees, indicating that *Drosophila* could be used as a model insect for genetic analysis of caste differentiation.

Royalactin changes *Drosophila* phenotypes via Egfr

The insulin signalling pathway in metazoans has an important role in regulating body size, growth and metabolism[15,16]. First, I examined the effects of royal jelly on body size of *insulin receptor* (*InR*) mutants (*InR*^{E19}/*InR*^{E19} and *InR*^{p5545}/*InR*^{E19})[15] and mutant showing elevated levels of phosphatidylinositol-3 kinase (PI3K) activity in prothoracic gland and corpora allata with *P0206-Gal4* (*P0206>dPI3K*)[16], all of which show reduced body size and weight. The *InR* mutants and *P0206>dPI3K* reared with royal jelly medium had larger body size and shorter developmental time than individuals reared with control medium or casein medium (Supplementary Fig. 6 and Supplementary Table 3).

I previously found that royalactin functions similarly to Egf in rat hepatocytes[11,12]. Therefore, I investigated the effects of royal jelly on body size of *Epidermal growth factor receptor* (*Egfr*) mutants (*Egfr*^{tsla}/ *Egfr*^{f24})[17]. Royal jelly did not influence body size or developmental time in the *Egfr* mutants (Supplementary Fig. 6 and Supplementary Table 3). Next, to determine the tissue specificity of royal jelly action in flies, I examined the influence of royal jelly on body size and developmental time in mutants in which expression of Egfr was silenced in the prothoracic gland, corpora allata or fat body, which are involved in body size regulation of *Drosophila*[16,18–21]. I used *Aug21-Gal4* (ref. 19) or *pumpless* (*ppl*)-*Gal4* (refs 16, 22) as a line with specific Gal4 expression in corpora allata or fat body, respectively. Royal jelly increased body size and shortened developmental time in *P0206>dEgfrRNAi*
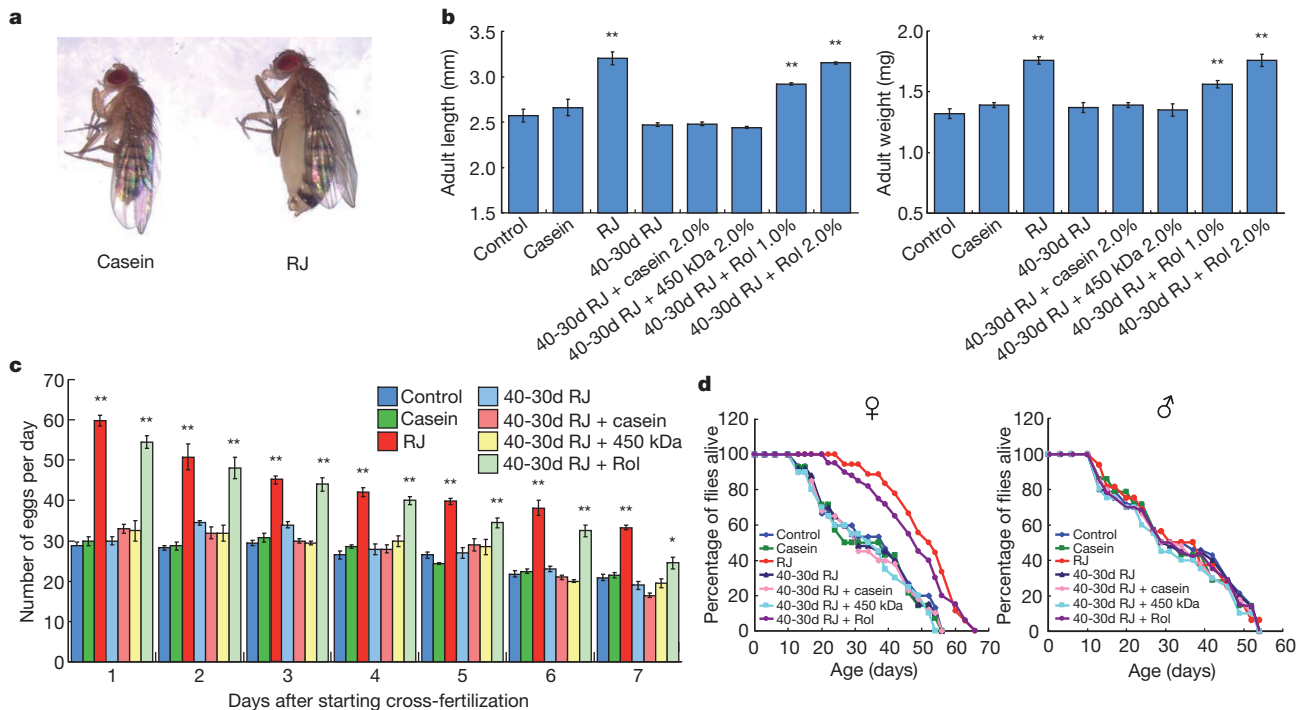
Figure 2 | Morphological and physiological changes of *Drosophila melanogaster* induced by royal jelly and royalactin. a, Body size of female adult flies reared with casein medium (8% yeast, 11.3% D-glucose, 2.8% casein, 1.3% D-fructose, 0.4% cornstarch, 0.76% soybean oil) and royal jelly medium (8% yeast, 10% D-glucose, 20% royal jelly). **b–d,** Body length (**b**, left), body weight (**b**, right), fecundity (**c**) and longevity (**d**) in wild-type (CS) fruit flies reared with control medium (8% yeast, 10% D-glucose), casein medium, royal jelly (RJ) medium or medium containing royal jelly stored at 40 °C for 30 days (40-30d RJ) or 40-30d RJ with casein, 450-kDa protein or royalactin (Rol) ($n > 40$). Values are expressed as mean ± s.e.m. Values significantly different from those of flies reared with control medium are indicated by *$P < 0.05$ or **$P < 0.01$.

flies (an RNA interference (RNAi) line for *Drosophila* Egfr) and *Aug21>dEgfrRNAi* flies, whereas it did not affect body size or developmental time in *ppl>dEgfrRNAi* flies (Supplementary Fig. 6 and Supplementary Table 3). Similar results were observed in *ppl>dEgfrRNAi* flies reared with royalactin (data not shown). I confirmed that *Drosophila* Egfr was expressed in the fat body of the wild-type flies (Supplementary Fig. 7). These findings demonstrated that not InR signalling, but rather Egfr signalling in the fat body was implicated in the increase of body size and reduction of developmental time by royal jelly or royalactin.

I next investigated how Egfr signals regulate changes of body size and developmental time in response to royal jelly. Royal jelly or royalactin activated S6K—which is activated by both phosphatidylinositol-dependent kinase 1 (PDK1) downstream of PI3K and target of rapamycin (TOR) downstream of PI3K/PDK1/Akt through stimulation of Egfr[23–26]—and mitogen-activated protein kinase (MAPK) in the larval fat body, and the activation of these enzymes by royalactin was suppressed by *Drosophila* Egfr RNAi in the fat body (Supplementary Fig. 8). Royal jelly did not increase the body size of *ppl>dPI3KDN* (*Drosophila* PI3K dominant-negative), *ppl>dPDK1RNAi*, *ppl>dAktRNAi*, *ppl>dTORDN* or *ppl>dS6KDN* flies, but shortened their developmental time, whereas *ppl>dRafRNAi* and *ppl>dMKP3* (ERK-inhibitory phosphatase)[27] reared with royal jelly showed increased body size but no early eclosion compared to the mutants reared with control medium or casein medium (Supplementary Fig. 6, Supplementary Table 3 and Supplementary Table 4). The increase of cell size in flies reared with royal jelly was repressed in *ppl>dEgfrRNAi* and *ppl>dS6KDN*, but not *ppl>dMKP3* (Supplementary Fig. 9). Loss of S6K function in *Drosophila* reduces body size by decreasing cell size but not cell number[28]. Activity of the MAPK pathway is reported to be unaffected by nutrients[29]. These results indicate that royalactin activated S6K through Egfr in the fat body, acting as a morphogenic factor to increase body size through an increase of cell size, and it also activated

the MAPK pathway in the fat body to reduce the developmental time in *Drosophila*.

Drosophila phenotypes change in response to royalactin overexpression

To examine the stimulatory action of royalactin on *Drosophila* further, I examined the effect of overexpression of *royalactin* using the UAS/Gal4 system[30]. Surprisingly, *act>royalactin* showed increased body size, cell size, fecundity and longevity and shortened developmental time compared with *UAS-royalactin* (Fig. 3b–d, Supplementary Fig. 10 and Supplementary Table 5). Moreover, overexpression of *royalactin* specifically in the fat body or an Egfr signal using ppl-Gal4 or Gal4 driver of rhomboid (rho), which is the essential signal-generating component of Egfr signalling during development in *Drosophila*[31], induced the same phenotypes as *act>royalactin* (Fig. 3, Supplementary Fig. 10 and Supplementary Table 5). Royal jelly proteins were reported to contain royalactin, identical to major royal jelly protein (MRJP)1 and MRJP2–5 (ref. 32). I overexpressed *mrjp2–5* with *act-Gal4*, *rho-Gal4* and *ppl-Gal4*, and found that the body sizes of these mutants overexpressing *mrjp2–5* did not change (Supplementary Table 6). Overexpression of royalactin activated MAPK and S6K in the fat body of larvae, and this activation was inhibited by *Drosophila* Egfr RNAi (Supplementary Fig. 11). On the other hand, when *royalactin* was overexpressed with *P0206-Gal4* or *Aug21-Gal4*, it did not influence body size or developmental time of the mutants (Fig. 3b and Supplementary Table 5). Increase of body size and cell size in *ppl>royalactin* and *rho>royalactin* was suppressed by inhibition of Egfr and S6K, but not by abrogation of InR and MAPK (Fig. 3b, Supplementary Fig. 10 and data not shown). Reduction of developmental time in *ppl>royalactin* and *rho>royalactin* was repressed by inhibition of Egfr and MAPK, but not by inhibition of S6K (Supplementary Table 5 and data not shown). These results are consistent with the findings in flies reared with royal jelly.
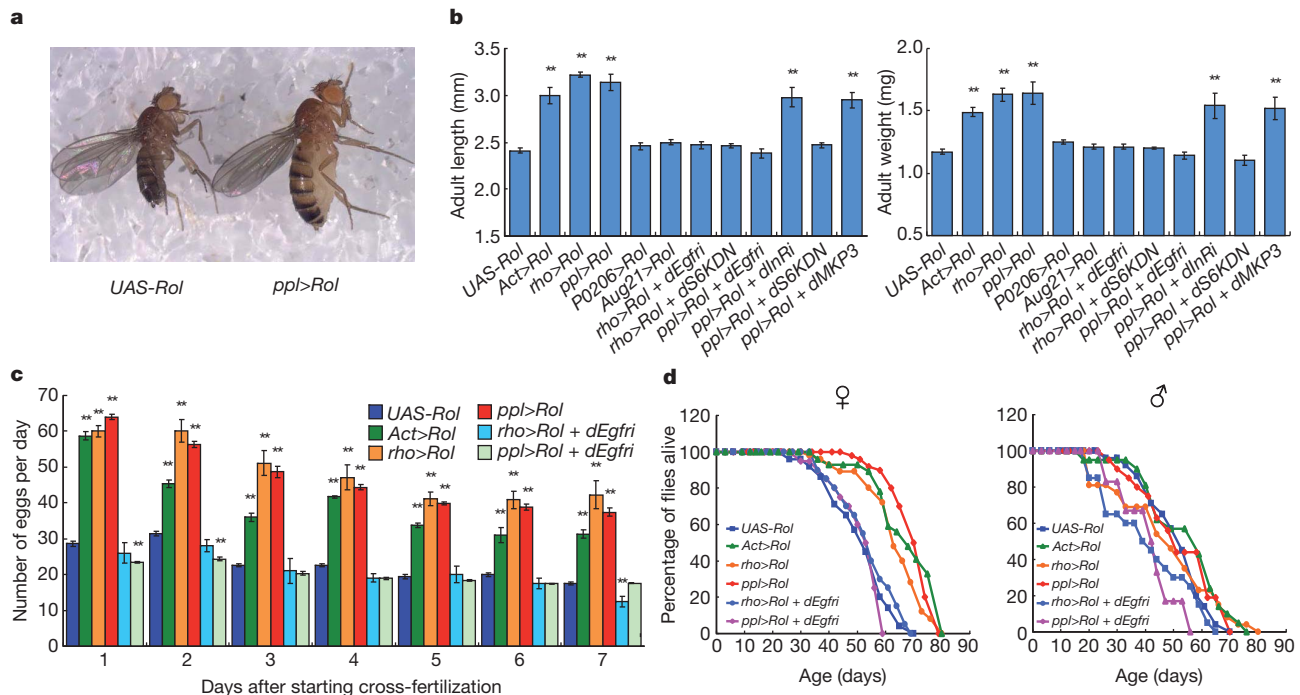
Figure 3 | Morphological and physiological changes of *Drosophila melanogaster* induced by overexpression of *royalactin*. a, Body size of female adult flies without or with overexpression of *royalactin* in the fat body (*UAS-Rol* or *ppl>Rol*). **b**, Body length (**b**, left) and body weight (**b**, right) in flies with overexpression of *royalactin* and in the signal factor suppression mutants in the *royalactin* overexpression background. *n* > 40. **c, d**, Fecundity (**c**) and longevity (**d**) in flies with overexpression of *royalactin* and in *Drosophila Egfr* interference (*dEgfri*) mutants in the *royalactin* overexpression background (*n* > 50). Values are expressed as mean ± s.e.m. Values significantly different from those of *UAS-royalactin* are indicated by **$P < 0.01$.

Royalactin changes hormone metabolism in *Drosophila*

To investigate the relationship between the morphological and physiological changes induced by royalactin in flies and hormone modulation, I measured changes in the biosynthesis of a biologically active ecdysteroid, 20-hydroxyecdysone (20E), and juvenile hormone in wild-type flies given royal jelly during the larval period. Moreover, changes in gene expression of *yolk protein* (*yp*) during larval development were examined because juvenile hormone induces expression in the fat body of *yp*, which is essential for vitellogenesis, thereby promoting egg production in *Drosophila*[33]. Royal jelly and royalactin increased the 20E titre at 3 days after egg deposition (AED), and juvenile hormone titre and gene expression of *yp* at 4 days AED (Supplementary Fig. 12 and Supplementary Fig. 13). The increase of 20E titre in flies reared with royal jelly was suppressed in *ppl>dEgfrRNAi* and *ppl>dMKP3*, but not *ppl>dS6KDN* (Supplementary Fig. 14a), indicating that activation of MAPK downstream of Egfr in the fat body by royalactin induced 20E synthesis to shorten the developmental time. On the other hand, the increase of juvenile hormone titre, gene expression of *yp* and fecundity by royal jelly was repressed in *ppl>dEgfrRNAi*, but not in *ppl>dS6KDN* or *ppl>dMKP3* flies (Supplementary Fig. 14b–d and Supplementary Fig. 15). Because repression of MAPK in the fat body (*ppl>dMKP3*) did not abrogate the increase of *yp* expression and fecundity, the increase of 20E by royalactin seemed not to be associated with the increase of *yp* expression and oviposition. Taken together, these findings indicated that Egfr signalling in the fat body is activated by royalactin via a pathway distinct from that regulating body size and developmental time, leading to induction of juvenile hormone synthesis and a consequent increase of *yp* expression, thereby increasing fecundity. S6K in the fat body also seemed be associated only with the increase of body size by royal jelly.

On the other hand, increase of fecundity in flies with overexpression of *royalactin* was also repressed by *Drosophila Egfr* RNAi in the fat body but not by suppression of S6K and MAPK in the fat body (Fig. 3c and data not shown). These results were consistent with those obtained in flies reared with royal jelly. Increase of longevity induced

by royal jelly was also abrogated in *ppl>dEgfrRNAi* flies, but not *ppl>dS6KDN* or *ppl>dMKP3* flies, indicating that Egfr in the fat body was essential for the increase of longevity in flies reared with royal jelly (Supplementary Fig. 14e and Supplementary Fig. 16a, b). Similar results were seen in the case of overexpression of *royalactin* (Fig. 3d and data not shown).

Suppression of queen differentiation in honeybees with RNAi

To confirm the signalling pathway involved in caste development, I reared honeybee larvae with suppression of *Apis mellifera InR* (*InR*) and *Egfr* by RNAi. Knockdown of *InR* did not affect final adult size, developmental time or ovary size in individuals reared with royal jelly, including a double-stranded RNA for green fluorescent protein (GFP), a control of RNAi, whereas *Egfr* RNAi reduced adult size and ovary size, and prolonged developmental time, compared with the control (GFP) (Fig. 4 and Supplementary Fig. 17a). These inhibitory effects of *Egfr* RNAi on queen differentiation were also observed in individuals reared with royalactin (data not shown). Royalactin activated MAPK and S6K through Egfr in fat body of honeybee larvae as effectively as did royal jelly (Supplementary Fig. 18). These results indicate that the activation of Egfr by royalactin is also involved in caste differentiation in the honeybee. Furthermore, suppression of honeybee *PI3K*, *PDK1*, *TOR* and *S6K* with RNAi inhibited the increase to final adult size induced by royal jelly, but did not affect changes of developmental time or ovary development (Fig. 4, Supplementary Fig. 17a and Supplementary Fig. 19). Royal jelly or royalactin increased the 20E titre in 3-day-old honeybee larvae, and the juvenile hormone titre and gene expression of *vitellogenin* (*vg*), a precursor of *yp*, in 4-day-old honeybee larvae given 40 °C/30 d royal jelly, whereas the 450-kDa protein and casein did not (Supplementary Fig. 4 and Supplementary Fig. 20). Increase of the 20E titre in honeybee larvae reared with royal jelly was abolished by *Egfr* RNAi and PD98059, a MAPK inhibitor, but not *S6K* RNAi (Supplementary Fig. 20a). PD98059 prolonged developmental time in larvae reared with royal jelly (data not shown). Increase
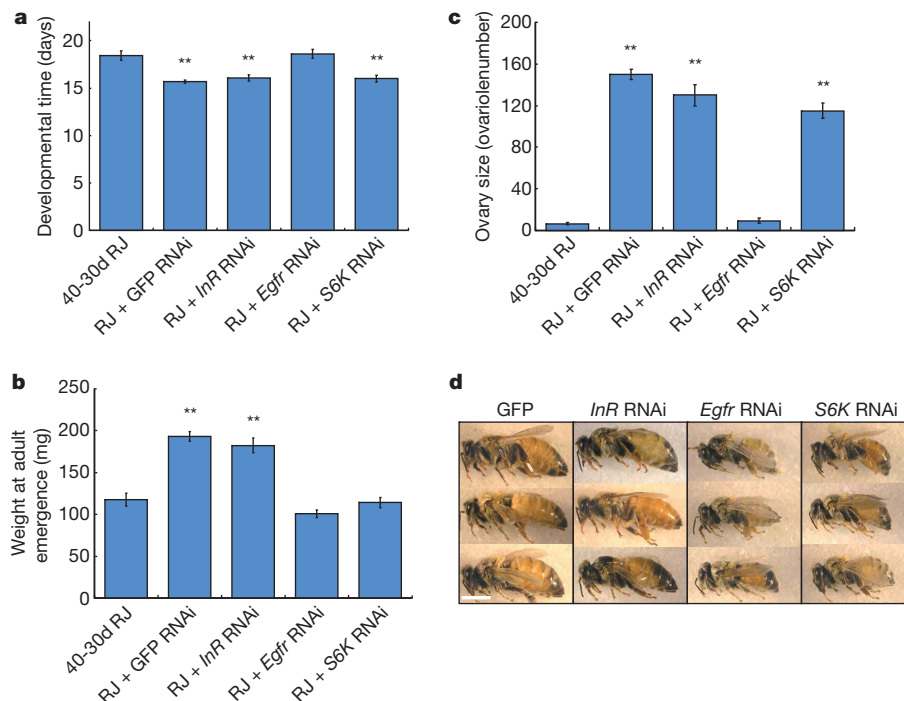
Figure 4 | Suppression of queen differentiation in honeybee with RNAi.
a–c, Developmental time (**a**), weight at adult emergence (**b**) and ovary size (**c**) in individuals ($n = 10$–16) reared with royal jelly stored at 40 °C for 30 days (40-30d RJ) and royal jelly (RJ) containing dsRNA of GFP and signal factors. Values are expressed as mean ± s.e.m. Values significantly different from those of larvae reared with 40-30d RJ are indicated by $**P < 0.01$. **d**, The final adult size after eclosion is shown in the photograph. Scale bar, 5 mm.

of juvenile hormone titre and *vg* expression in honeybee larvae reared with royal jelly was inhibited by *Egfr* RNAi but not *S6K* RNAi and PD98059 (Supplementary Fig. 20b, c). Thus, I found that activation of S6K by royalactin through Egfr was involved in the increase of body size in queens, whereas MAPK activity downstream of Egfr in response to royalactin was responsible for the increase of 20E synthesis, thereby shortening developmental time in the honeybee. Topical application of juvenile hormone to worker larvae results in the emergence of queen-like individuals with ovary development, but which display body sizes consistent with workers[34,35]. Therefore, an increase in juvenile hormone titre downstream of Egfr signalling activated by royalactin may have a function in ovary development in queens. These mechanisms are consistent with those of the morphological and physiological changes induced by royalactin in flies.

Here I provide the first evidence, to my knowledge, that royalactin acts on Egfr in the honeybee to induce queen differentiation. Furthermore, I found that administration and overexpression of royalactin in *Drosophila* caused morphological and physiological changes resembling the phenotypes of queen bee, through a similar mechanism to that of caste differentiation in the honeybee. These results provide new evidence that Egfr signalling has an important role in growth regulation.

The 450-kDa protein consists of apalbumin (420 kDa) and an oligomer of apisimin (5.5 kDa)[36]. Apalbumin is an oligomer of the gene product of *mrjp1* (ref. 36). On the other hand, royalactin is derived from *mrjp1* (ref. 37), but is present as a monomeric glycoprotein with a molecular mass of 57 kDa in royal jelly, and is structurally distinct from apalbumin (antibodies to royalactin do not recognize apalbumin)[11]. Apalbumin binds strongly to apisimin to form a stable complex (450-kDa protein); apalbumin was not separated from apisimin in the absence of detergent[36]. These results indicate that royalactin is not derived from apalbumin in royal jelly. The 40 °C/30 d royal jelly, which contained 90% of the initial concentration of 450-kDa protein, did not induce queen development, and the 450-kDa protein did not increase the rate of emergence of queens when it was added to a diet containing 40 °C/30 d royal jelly. However, both royalactin and E-royalactin

induced queen differentiation in the honeybee. Thus, only royalactin, a monomer of MRJP1, functions as a caste determination factor. Royalactin induced prolonged longevity through Egfr in *Drosophila*, indicating that royalactin might have an important role in the prolongation of longevity in queens. To my knowledge, this is the first evidence that Egfr is involved in the regulation of longevity. Further research will be required to investigate the mechanism through which royalactin regulates lifespan in the fruitfly and the honeybee.

The association between royal jelly and caste formation has been known for more than 100 years, but the identity of the component(s) in royal jelly that induces queen development has been elusive. My results provide important insights into the process of caste development in the honeybee, and may also offer a valuable clue to eusociality and the evolution of social hymenopterans.

METHODS SUMMARY

Fly larvae were reared with medium containing royal jelly, D-glucose, yeast and agar at 25 °C. Honeybee larvae were reared with medium containing royal jelly, D-glucose, D-fructose and yeast extract at 34 °C with 96% humidity. Quantitative assay of juvenile hormone was carried out by high-resolution liquid chromatography-mass spectrometry (LC-MS) on a micrOTOF-Q instrument. The 20E titre of larvae was determined by the enzyme immunoassay (EIA) method. Quantitative analysis of gene expression was conducted by real-time PCR with the primers shown in Supplementary Tables 12 and 13. For honeybee RNAi experiments, the rearing diet containing enzymatically synthesized dsRNA at $150\,\mu g\,ml^{-1}$ was administered to second instar larvae for 2 days.

1. Maynard Smith, J. & Szathmary, L. *The Major Transitions in Evolution* (Freeman, 1995).
2. Haydak, M. H. Honey bee nutrition. *Annu. Rev. Entomol.* **15,** 143–156 (1970).
3. Patel, N. G., Haydak, M. H. & Gochnauer, T. A. Electrophoretic components of the proteins in honeybee larval food. *Nature* **186,** 633–634 (1960).
4. Weaver, N. Effects of larval age on dimorphic differentiation of the female honey bee. *Ann. Entomol. Soc. Am.* **50,** 283–294 (1957).
5. Shuel, R. W. & Dixon, S. E. The early establishment of dimorphism in the female honeybee, *Apis mellifera* L. *Insectes Soc.* **7,** 265–282 (1960).

6. Page, R. E. & Peng, C. Y. Aging and development in social insects with emphasis on the honey bee, *Apis mellifera* L. *Exp. Gerontol.* **36,** 695–711 (2001).
7. Wheeler, D. E. Developmental and physiological determinations of caste in social Hymenoptera: evolutionary implications. *Am. Nat.* **128,** 13–34 (1986).
8. Patel, A. *et al.* The making of a queen: TOR pathway is a key player in diphenic caste development. *PLoS ONE* **2,** e509 (2007).
9. Kamakura, M., Mitani, N., Fukuda, T. & Fukushima, M. Antifatigue effect of fresh royal jelly in mice. *J. Nutr. Sci. Vitaminol. (Tokyo)* **47,** 394–401 (2001).
10. Beetsma, J. The process of queen-worker differentiation in the honeybee. *Bee World* **60,** 24–39 (1979).
11. Kamakura, M., Suenobu, N. & Fukushima, M. 57-kDa protein in royal jelly enhances proliferation of primary cultured rat hepatocytes and increases albumin production in the absence of serum. *Biochem. Biophys. Res. Commun.* **282,** 865–874 (2001).
12. Kamakura, M. Signal transduction mechanism leading to enhanced proliferation of primary cultured adult rat hepatocytes treated with royal jelly 57-kDa protein. *J. Biochem.* **132,** 911–919 (2002).
13. Bloch, G., Wheeler, D. E. & Robinson, G. E. in *Hormones, Brain and Behavior* Vol. 3 (eds Pfaff, D. W., Arnold, A. P., Fahrbach, S. E., Etgen, A. M. & Rubin, R. T.) 195–235 (Academic Press, 2002).
14. Wirtz, P. & Beetsma, J. Induction of caste differentiation in the honeybee (*Apis mellifera* L.) by juvenile hormone. *Entomol. Exp. Appl.* **15,** 517–520 (1972).
15. Tatar, M. *et al.* A mutant *Drosophila* insulin receptor homolog that extends life-span and impairs neuroendocrine function. *Science* **292,** 107–110 (2001).
16. Colombani, J. *et al.* Antagonistic actions of ecdysone and insulins determine final size in *Drosophila. Science* **310,** 667–670 (2005).
17. Gilboa, L. & Lehmann, R. Soma-germline interactions coordinate homeostasis and growth in the *Drosophila* gonad. *Nature* **443,** 97–100 (2006).
18. Caldwell, P. E., Walkiewicz, M. & Stern, M. Ras activity in the *Drosophila* prothoracic gland regulates body size and developmental rate via ecdysone release. *Curr. Biol.* **15,** 1785–1795 (2005).
19. Mirth, C., Truman, J. W. & Riddiford, L. M. The role of the prothoracic gland in determining critical weight for metamorphosis in *Drosophila melanogaster. Curr. Biol.* **15,** 1796–1807 (2005).
20. Belgacem, Y. H. & Martin, J. R. Hmgcr in the *Corpus Allatum* controls sexual dimorphism of locomotor activity and body size via the insulin pathway in *Drosophila. PLoS ONE* **2,** e187 (2007).
21. Colombani, J. *et al.* A nutrient sensor mechanism controls *Drosophila* growth. *Cell* **114,** 739–749 (2003).
22. Zinke, I. *et al.* Suppression of food intake and growth by amino acids in *Drosophila*: the role of *pumpless*, a fat body expressed gene with homology to vertebrate glycine cleavage system. *Development* **126,** 5275–5284 (1999).
23. Navolanic, P. M., Steelman, L. S. & McCubrey, J. A. EGFR family signaling and its association with breast cancer development and resistance to chemotherapy. *Int. J. Oncol.* **22,** 237–252 (2003).
24. LeVea, C. M., Reeder, J. E. & Mooney, R. A. EGF-dependent cell cycle progression is controlled by density-dependent regulation of Akt activation. *Exp. Cell Res.* **297,** 272–284 (2004).
25. Rintelen, F., Stocker, H., Thomas, G. & Hafen, E. PDK1 regulates growth through Akt and S6K in *Drosophila. Proc. Natl Acad. Sci. USA* **98,** 15020–15025 (2001).
26. McManus, E. J. & Alessi, D. R. TSC1–TSC2: a complex tale of PKB-mediated S6K regulation. *Nature Cell Biol.* **4,** E214–E216 (2002).
27. Lee, K. S. *et al. Drosophila* short neuropeptide F signalling regulates growth by ERK-mediated insulin signalling. *Nature Cell Biol.* **10,** 468–475 (2008).
28. Montagne, J. *et al. Drosophila* S6 kinase: a regulator of cell size. *Science* **285,** 2126–2129 (1999).
29. Zhang, H. *et al.* Regulation of cellular growth by the *Drosophila* target of rapamycin dTOR. *Genes Dev.* **14,** 2712–2724 (2000).
30. Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118,** 401–415 (1993).
31. Urban, S. Rhomboid proteins: conserved membrane proteases with divergent biological functions. *Genes Dev.* **20,** 3054–3068 (2006).
32. Schmitzova, J. *et al.* A family of royal jelly proteins of the honeybee *Apis mellifera L. Cell. Mol. Life Sci.* **54,** 1020–1030 (1998).
33. Bownes, M. The regulation of the yolk protein genes, a family of sex differentiation genes in *Drosophila melanogaster. Bioessays* **16,** 745–752 (1994).
34. Goewie, E. A. & Beetsma, J. Induction of caste differentiation in the honey bee (*Apis mellifera* L.) after topical application of JH-III. *Proc. K. Ned. Akad. Wet. C* **79,** 466–469 (1976).
35. Ebert, R. Influence of juvenile hormone on gravity orientation in the female honeybee larvae (*Apis mellifera* L.). *J. Comp. Physiol.* **137,** 7–16 (1980).
36. Bílikova, K. *et al.* Apisimin, a new serine-valine-rich peptide from honeybee (*Apis mellifera* L.) royal jelly: purification and molecular characterization. *FEBS Lett.* **528,** 125–129 (2002).
37. Kamakura, M. & Sakaki, T. A hypopharyngeal gland protein of the worker honeybee *Apis mellifera* L. enhances proliferation of primary-cultured rat hepatocytes and suppresses apoptosis in the absence of serum. *Protein Expr. Purif.* **45,** 307–314 (2006).